



Créer, trouver et exploiter les blogs

Olivier Ertzscheid

► To cite this version:

| Olivier Ertzscheid. Créer, trouver et exploiter les blogs. ADBS éditions, 64 p., 2008. sic_00586027v2

HAL Id: sic_00586027

https://archivesic.ccsd.cnrs.fr/sic_00586027v2

Submitted on 21 Apr 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Olivier Ertzscheid

Créer, trouver et exploiter les blogs



À NTM...
(Noa, Tom & Milo)
mes trois petits bonshommes,
en attendant qu'ils ouvrent leur blog
O.E.

Dans la même collection

« L'Essentiel sur... »

Utiliser les fils RSS et ATOM

Par Serge Courrier. 2007

Documentalistes, ajoutez de la valeur à vos services !

Par Jacques Chaumier et Éric Sutter. 2007

L'apport des technologies linguistiques au traitement et à la valorisation de l'information textuelle

Par Bernard Normier. 2007

Le panorama de presse : aspects juridiques

Par Michèle Battisti. 2006

L'évaluation et les indicateurs de la performance des activités info-documentaires

Par Éric Sutter. 2006

Les dossiers documentaires : dossier outil, dossier produit, dossier électronique

Par Clotilde Vaissaire et Françoise Quaire, en collaboration avec Bruno Salléras. 2005

Le management de l'information : présentation commentée du document de normalisation X50-185

Par Éric Sutter. 2005

Vade-mecum des chercheurs d'images : petit guide pratique à l'usage des iconographes et des chercheurs

2^e édition coordonnée par Isabelle Jullien et Marie-Odile Perulli (secteur Audiovisuel de l'ADBS) en collaboration avec l'Association nationale des iconographes (ANI). 2005

Sommaire

Introduction	5
« J'ai dix ans »	5
Le pari de l'instant	6
Premiers éléments contextuels	7
Le temps des pionniers (1995-1997)	7
Le temps des <i>early-adopters</i> (1997-2000)	8
Un temps pour tous (2000-...)	8
Quelques dates et quelques chiffres	9
Chapitre 1 - Comment ça marche ?	10
Qu'est-ce qu'un blog ?	10
Blogs : les raisons du succès	11
Créer son blog	20
Quelques blogs	22
Chapitre 2 - Où trouver des blogs et qu'en faire ?	26
Combien de blogs ?	26
Une influence certaine et une légitimité (partiellement) acquise	27
Trouver des blogs hier...	28
Trouver des blogs aujourd'hui...	30
Trouver des blogs demain...	40
Blogs : demain... la veille ?	41
Chapitre 3 - Quelle typologie pour quels contenus ?	48
Quelle typologie ?	48
Petite sociologie du blogueur	52
Le profil-type du chercheur de blogs	52
Approfondir...	53
Exemples et usages	56

Conclusion « J'ai dix ans »... de plus	.61
De quoi seront faits les blogs dans dix ans?	.61
La généralisation des blogs est effective	.61
Leur banalisation est actée	.62
Annexes	.63
Références pour approfondir...	.63
Compléments en ligne	.64

Introduction

« J'ai dix ans »



C'est le 17 décembre 1997 que Jorn Barger a contracté les mots « web » et « log » en « weblog ».

Les blogs occupent aujourd'hui une place centrale dans notre accès quotidien à l'information. Véritable écosystème informationnel, la « blogosphère¹ » est largement protéiforme. Du côté des individualités, les blogs d'adolescents y côtoient les blogs de chercheurs, de citoyens, d'hommes politiques, de journalistes... Du côté des collectifs, organismes, institutions, universités, entreprises et associations ont également « pignon sur blogosphère ».

« LA » blogosphère? Les moteurs de recherche en ont fait une source d'information comme une autre, mais parfois beaucoup plus présente ou représentée que certaines autres. Dans « la » blogosphère, la presse puise certaines informations, certains témoignages; dans « la » blogosphère, les entreprises tiennent boutique et font commerce; dans « la » blogosphère, les universitaires enseignent, chroniquent, analysent; les adolescents s'y épanchent, s'y dévoilent, s'y racontent; les cuisinières livrent leurs recettes; les avocats plaident; les *geeks*² codent; les dessinateurs de BD dessinent; les politiques font campagne. Pas de métier, de champ professionnel, de compétence, de loisir, de passion qui n'y échappe.

1 Le terme « blogosphère » désigne l'ensemble des blogs publics et les relations qu'ils entretiennent entre eux.

2 Terme anglais désignant une personne passionnée, voire obsédée, par l'informatique et ses applications.

◆ Le pari de l'instant

En écho aux constats ci-dessus, la première question à laquelle entend répondre cet ouvrage est donc : « Pourquoi ? » Il faudra ensuite comprendre « comment ». Comment s'y repérer, s'y orienter, y flâner sans s'y perdre ? Comment utiliser au mieux les outils de recherche pour dénicher la perle rare ou mener une recherche un tant soit peu systématique ?

Une fois ces questions éclaircies, il nous restera à proposer une grille de lecture, une typologie pour mieux comprendre et caractériser ces contenus, pour tenter d'englober cette diversité, de la rassembler sur une carte, et de mettre cette carte sous vos yeux. Tout en sachant qu'aucune carte ne reflète jamais complètement la richesse du paysage.

La principale difficulté de cet ouvrage est la rapidité. Rapidité du format de cette collection tout d'abord. Rapidité technologique ensuite, au nom de laquelle peu d'outils ont le temps de durer – tant la dernière version remplace promptement la précédente – et peu d'usages le temps de se fixer. Les blogs n'échappent pas à cette règle. L'avenir des blogs, de leur format, de leurs usages, sera très probablement riche en surprises, en bouleversements, en ruptures... et en continuités nouvelles.

Veuillez donc considérer cet ouvrage comme celui d'un instant. Construit sur ceux qui l'ont précédé et tentant d'annoncer ceux qui le suivront. Il s'adresse d'abord à l'« apprenant », au novice, au candide, au curieux, à l'Homme pressé. S'ils peuvent y puiser de quoi emboîter les pas de l'expert, du *geek*, de l'« habitué » et si ces trois derniers n'y trouvent pas grand-chose à redire, se surprenant même à trouver parfois cette compagnie agréable, stimulante, efficace, alors cet ouvrage aura atteint son but.

Bonne lecture donc. Et bon blog 😊.

Premiers éléments contextuels



Après l'avènement de l'hypertexte et la naissance conjointe du « web », après celui du courriel et des listes et groupes de discussion, après l'essor et l'engouement relatif aux technologies de la messagerie instantanée, et conjointement au développement des technologies WiFi et de l'internet sans fil, les « weblogs », « blogs », « bloc-notes », « blogues » ou « carnets web » sont l'une des dernières évolutions marquantes liées aux technologies de l'information et de la communication.

◆ Le temps des pionniers (1995-1997)

Hier les premiers blogs³ étaient constitués de simples listes de liens pointant vers de nouveaux sites, parfois classés par date, parfois agrémentés de succincts commentaires de l'auteur desdites compilations.

Quelques repères

- Tim Berners Lee alimente dès 1992 une rubrique « What's new » sur son site personnel. Certains considèrent cela comme le tout premier blog.
- Dave Winer en avril 1997 démarre un blog⁴ (liens commentés par l'auteur) au style plus « journalistique ». Toujours actif, ce blog est sans conteste celui qui a la plus longue durée de vie sans interruption sur Internet.
- En septembre 1997, Rob Malda démarre SlashDot⁵, introduit le concept de commentaires des utilisateurs et rencontre une très forte popularité.
- En décembre 1997 John Barger utilise le terme « weblog » dans son journal en ligne⁶.

3 Pour un historique plus complet que les quelques éléments mentionnés ici, on se reportera à : Rebecca Blood, *Weblogs: A History and Perspective* (voir la référence page 63). On consultera également la page Weblog du Meatball Wiki : www.usemod.com/cgi-bin/mb.pl?CarnetWeb

4 www.scriptingnews.com

5 <http://slashdot.org>

6 www.robotwisdom.com

◆ Le temps des *early-adopters* (1997-2000)

L'arrivée sur le marché (américain) d'outils extrêmement simples de mise en forme et d'organisation va rapidement propulser le blog au rang « d'application grand public ». La gratuité de ces mêmes outils fera le reste.

En 1997 apparaissent les premiers outils de gestion de contenu. Ils ont pour noms Pitas⁷, Groupsoup, Metafilter, Squishdot ou encore... Blogger. Ils permettent essentiellement d'entrer l'adresse (URL) du site référencé, d'affecter un titre à chaque nouvelle entrée, et de déposer des commentaires.

En 1999, plusieurs services de publication automatique de weblogs sont disponibles, dont le célèbre LiveJournal⁸, lancé en mars 1999 par Brad Fitzpatrick. Leur utilisation se simplifie encore. La gestion et la publication d'un blog deviennent alors plus simples, rapides et pratiques que celles d'un site web personnel (qui continue de nécessiter la maîtrise d'un code informatique, le langage HTML). La – bonne – nouvelle ne tardera pas à se répandre...

Le 23 août 1999, la société PyraLabs fonde et lance le service Blogger. Quatre ans plus tard, le 14 février 2003, il sera racheté par le moteur Google. Comptant 1,1 million d'utilisateurs à la date de l'achat, il reste aujourd'hui l'un des services de *blogging* les plus populaires.

Le 17 avril 2000, *The Guardian* démarre son propre blog⁹. C'est la première plongée d'un grand média dans la blogosphère.

◆ Un temps pour tous (2000-...)

Un blog est une page personnelle, généralement alimentée par un seul individu et constituée de courts messages affichés et archivés automatiquement selon un ordre ante-chronologique. Mais plusieurs déclinaisons de cette forme originelle sont possibles¹⁰ : ainsi il existe des weblogs collaboratifs entre plusieurs auteurs, des weblogs d'entreprise qui suppléent à toute la chaîne de gestion de la relation-client (CRM), font office d'intranets, absorbent les tâches liées à la gestion de projet, au management des connaissances, etc.

C'est en 2002 qu'un blog se trouve pour la première fois propulsé à la une de tous les médias de la planète : lors de la guerre en Irak, la seule source d'information « locale » est le blog de celui qui se fait appeler Salam Pax et que la presse rebaptise « le blogueur de Bagdad ».

7 Fondé par Andrew Zeepo, le service Pitas fut l'un des premiers à offrir un gabarit automatique en HTML : www.pitas.com

8 www.livejournal.com

9 www.guardian.co.uk/weblog

10 [www.google.com/search?hl=en&lr=&q=define %3Ablog&btnG=Search](http://www.google.com/search?hl=en&lr=&q=define+%3Ablog&btnG=Search)

◆ Quelques dates et quelques chiffres

- Janvier 1999 : 23 blogs sont répertoriés¹¹.
- Septembre 2003 : le moteur Technorati¹² recense 4,1 millions de blogs.
- 4 octobre 2003 : première conférence internationale sur les blogs à l'université de Harvard¹³.
- Octobre 2004 : Technorati recense 4 millions de blogs.
- Décembre 2004 : Technorati recense 6,5 millions de blogs.
- Mars 2005 : Technorati recense 7,8 millions de blogs.
- Avril 2007 : Technorati recense 70 millions de blogs.

Ces chiffres, s'ils confirment l'ampleur du phénomène et sa très large adoption, doivent cependant être nuancés sur deux points.

Primo, le nombre de blogs existants n'a que peu à voir avec le nombre de blogs « actifs ». Cette dernière notion est d'ailleurs elle-même délicate à définir et les critères varient selon les études (« l'activité » d'un blog nécessite au moins une mise à jour tous les trois mois pour certaines, une mise à jour par mois pour d'autres). Il est en tout cas établi qu'au moins la moitié des blogs existants sont inactifs ou moribonds¹⁴ et qu'ils sont en tout cas bien loin de l'idéal théorique d'une mise à jour quotidienne ou même hebdomadaire.

Secundo, quels que soient les chiffres annoncés, aucune étude ne recense la totalité des weblogs existants, le nombre des plateformes et solutions d'hébergement étant lui-même délicat à inventorier de manière fiable¹⁵.

Néanmoins, un certain nombre d'indicateurs et des retours d'expérience de plus en plus nombreux permettent de comparer l'engouement pour leur utilisation à celui rencontré il y a quelques années pour l'usage du courrier électronique, et leur taux de croissance à celui du Web à ses débuts.

11 www.jjg.net/portal/tpoowl.html : une des toutes premières pages à tenter d'inventorier et de classer les blogs existants à l'époque.

12 www.technorati.com

13 <http://web.archive.org/web/20030605083152/http://blogs.law.harvard.edu/bloggerCon/>

14 Sur les 4.1 millions de l'étude Perseus, seuls 1.4 million étaient actifs, c'est-à-dire mis à jour au moins une fois tous les trois mois.

15 Ainsi certaines communautés, telle « Skyblogs » (www.skyblog.com/), service de création et d'hébergement gratuit offert aux auditeurs de la radio Skyrock, ont rapidement atteint le chiffre du million de blogs créés.

Chapitre 1

Comment ça marche ?



◆ Qu'est-ce qu'un blog ?

Une définition (parmi d'autres)

Le site Pointblog.com¹⁶ fut l'un des premiers à définir de manière simple et consensuelle ce qu'est un blog : « De web – toile – et log – carnet de bord –, un blog est un site web (...) composé essentiellement d'actualités, publiées au fil de l'eau et apparaissant selon un ordre ante-chronologique, le plus souvent enrichies de liens externes et de commentaires. »

L'encyclopédie collaborative Wikipédia reprend cette définition en la développant : « Un blog ou blogue (mot-valise de web log) est un site web constitué par la réunion de billets écrits dans l'ordre chronologique, et classés la plupart du temps par ordre ante-chronologique (les plus récents en premiers). Chaque billet (appelé aussi note ou article) est, à l'image d'un journal de bord ou d'un journal intime, un ajout au blog ; le blogueur (celui qui tient le blog) y délivre un contenu souvent textuel, enrichi d'hyperliens et d'éléments multimédias, sur lequel chaque lecteur peut généralement apporter des commentaires ou opinions personnelles.¹⁷ »

Enfin, la Commission générale de terminologie et de néologie¹⁸ dans le *Journal officiel* 20 mai 2005 recommande de franciser le terme en « bloc-notes »¹⁹.

En 2005, le mot « blog » fait officiellement son entrée dans le dictionnaire Merriam-Webster²⁰. Il est aujourd'hui présent dans l'ensemble des dictionnaires et encyclopédies de la planète.

¹⁶ www.lemondedublog.com/abc/

¹⁷ <http://fr.wikipedia.org/wiki/Blog>

¹⁸ Cette commission est rattachée au ministère de la Culture et de la Communication.

¹⁹ Définition accessible sur le site FranceTerme (<http://franceterme.culture.fr>)

²⁰ www.merriam-webster.com/

Si les blogs font aujourd'hui office tout à la fois de pages personnelles et de journaux de bord ou de journaux intimes, la difficulté (et la richesse...) de ces publications est de n'être réductible à aucun de ces deux critères : comme nous le verrons dans cet ouvrage, les blogs sont bien plus que de simples pages personnelles et ils ne sont pas « simplement » des journaux intimes.

Un auteur... des auteurs. Des contenus... des supports

Sur un blog chacun devient, en un instant et sans barrière « technique », auteur et producteur de contenus. Essentiellement limités au format texte lors de l'explosion grand public du phénomène, ces contenus sont aujourd'hui résolument multi-supports (on parle ainsi de vidéo-blogs, photo-blogs, audio-blogs).

Si la majorité des blogs ne comptent qu'un seul auteur, on compte aujourd'hui un grand nombre de blogs multi-auteurs, notamment sur des sites de presse ou dans un contexte associatif.

Ces derniers temps, on a vu émerger le concept de *blogsitting* et *blog-sitters* (littéralement : « baby-sitters de blogs »), pour permettre aux blogueurs devant s'absenter ou prendre des vacances de ne pas laisser leur blog « sans activité ». Des services spécialisés permettent de trouver la/le *blog-sitter* idéal²¹. Au-delà de leur côté amusant (ou déconcertant pour le novice), ces services témoignent de la relation particulière qui unit un auteur à son blog, ainsi que de l'importance cruciale de la régularité dans la publication des billets.

◆ Blogs : les raisons du succès

Une plateforme « clé en main », rapide, facile, gratuite

Plusieurs facteurs convergents expliquent l'incroyable succès que connaissent les blogs dans toutes les sphères informationnelles (pages personnelles, sites d'entreprise, intranets, chroniques journalistiques, etc.).

Premier de ces facteurs : la gratuité et l'apparition sur le marché, dès juillet 1999, d'outils grand public gratuits et faciles d'emploi avec lesquels il n'est plus nécessaire d'avoir une connaissance même minimale d'HTML²²

²¹ www.blogsitter.net/

²² HTML : HyperText Markup Language (langage de balisage hypertexte) : c'est le langage qui permet de créer des pages web.

pour publier des pages sur le Net. Le plus utilisé de ces outils est Blogger²³, racheté par Google en février 2003.

Vient ensuite la « facilité » et la transparence du processus de publication : la mise en ligne de contenus se fait de manière simple et intuitive. Quelle que soit la plateforme choisie, tout se déroule à partir d'une interface WYSIWYG²⁴. Il n'est techniquement pas plus compliqué de poster un billet sur un blog que d'envoyer un courrier électronique.

Autre facteur clé de succès, la gestion automatique de l'information qu'auto-risent ces plateformes, qui s'organise autour de cinq particularités :

- les informations « postées » sont automatiquement archivées, l'utilisateur pouvant d'un simple clic choisir et modifier la fréquence de cet archivage (mensuel, hebdomadaire, quotidien) ;
- chaque information se voit automatiquement attribuer un URL permanent (permalien) qui permet de résoudre l'un des problèmes majeurs du web : celui de l'adressage et des liens brisés (la fameuse « Erreur 404 »). Quelle que soit la place (catégorie) à laquelle est publié un billet, et même si celle-ci est modifiée au cours de l'évolution du blog, son permalien restera, lui, inchangé ;
- chaque information est « indexée » à l'aide de métadonnées minimales : auteur, date de publication, titre, catégorie. Chacune de ces métadonnées peut ainsi servir de « filtre » lors de la navigation ou d'une recherche d'information ;
- chaque information peut être complétée *via* différentes fonctionnalités de nature collaborative : il est ainsi possible de poster des commentaires (liés au message et également archivés) et, dans le cas de blogs multi-auteurs, d'ouvrir des droits différents en lecture et en écriture pour chacun d'entre eux, selon différents statuts²⁵ ;
- les principales plateformes de publication proposent l'installation gratuite d'un moteur de recherche spécifique au blog, permettant une entrée « plein texte » sur l'ensemble de son contenu.

Comme le résume rapidement cet inventaire non exhaustif, les weblogs offrent ainsi au plus grand nombre les attendus techniques minimaux de tout système d'information « élaboré » : indexation, gestion des droits, archivage, recherche. C'est d'ailleurs à ce titre qu'ils remplacent d'ores et déjà dans cer-

²³ www.blogger.com

²⁴ WYSIWYG : acronyme de What You See Is What You Get. Désigne une interface qui ne nécessite pas de taper de code informatique. Toutes les mises en forme se présentent sous forme de « boutons », à la manière d'un traitement de texte.

²⁵ La plateforme Wordpress propose 4 statuts différents : l'administrateur (il peut tout modifier sur le blog), l'éditeur (il peut tout faire sauf créer de nouveaux utilisateurs et supprimer le blog), l'auteur (il peut modifier et publier ses propres billets uniquement), et le simple contributeur (il peut éditer ses billets mais ne peut pas les publier). <http://faq.wordpress.com/2006/05/12/what-are-the-different-roles/>

tains grands groupes les architectures lourdes et trop souvent sous-utilisées de type intranet.

Des fonctionnalités spécifiques

Si l'on regarde les plateformes de création de blogs, plusieurs fonctionnalités permettent de caractériser ce format éditorial. Pour bien comprendre la logique de l'outil, il faut se le représenter à la manière d'un puzzle dont chacune des différentes pièces est entièrement paramétrable et positionnable dans la structure d'ensemble.

Les billets

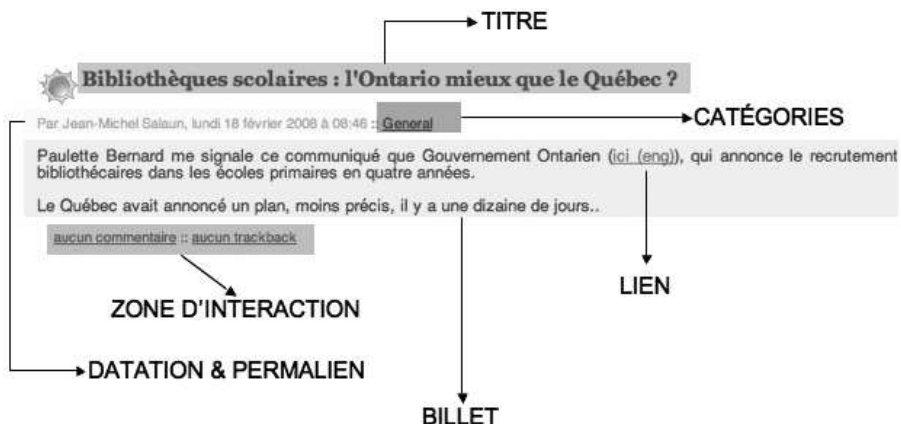
Également appelés « posts » ou « entrées », les billets constituent l'unité minimale de publication du blog. Chaque billet peut bénéficier des fonctionnalités décrites ci-après (commentaires, archivage, catégorisation, etc.).

On a souvent parlé à propos des blogs de « nanopublication » pour bien distinguer les « billets de blogs » des « pages de sites web ». Pour autant et à l'instar de ces dernières, tous les billets sont autant d'entrées possibles sur le weblog.

Chaque billet présente l'avantage d'être repérable *via* ce qui correspond à autant de métadonnées (son titre, son auteur, sa date de publication, ses mots clés, etc.). Comme une page web, un billet peut inclure des liens, de la vidéo, des images, du son.

Selon les plateformes de blogs, le permalien (adresse permanente du billet) est associé (« cliquable ») avec la zone de datation, le titre du billet, ou les deux.

FIGURE 1 – ANATOMIE D'UN BILLET DE BLOG



À noter. Les billets peuvent être affichés différemment selon les préférences de l'auteur du blog et en fonction de la manière dont les lecteurs y accèdent (par leur navigateur web ou *via* un lecteur RSS²⁶) : l'auteur peut décider de n'afficher que le titre du billet, un court extrait, ou le billet en intégralité. Ces « techniques », qui relèvent d'un simple paramétrage, sont notamment utilisées par les blogs à forte audience : en « obligeant » le lecteur à cliquer sur le billet pour en lire la suite, ou à laisser de côté son agrégateur pour afficher le blog dans un navigateur, il est ainsi possible de drainer du trafic supplémentaire ou de rentabiliser les visites *via* l'affichage de publicités.

Les commentaires

Chaque billet offre la possibilité de déposer automatiquement des commentaires. Ceux-ci peuvent à tout moment être fermés, modérés *a priori* ou restreints à une liste d'utilisateurs autorisés. Ce paramétrage peut s'appliquer à un billet particulier ou à l'ensemble des billets du blog. Il est ainsi possible de « fermer » les commentaires sur d'anciens billets (pour éviter le spam) tout en les maintenant actifs sur les autres. Enfin, il est possible de « programmer » la date de fermeture des commentaires, ou bien d'obliger les commentateurs à s'identifier en laissant leurs nom et adresse mail (qui peuvent être laissés visibles pour tous, ou seulement pour l'administrateur du blog).

Quel intérêt par rapport aux forums et autres babillards ? Le dépôt de commentaires sur un blog est moins complexe que sur certains forums. En outre, il s'agit d'une interaction plus « adressée » et contextuelle. Le commentaire est effectué non pas par rapport à un simple *thread* (thème ou fil de discussion) mais par rapport à un contenu plus complet. Même s'il n'existe pas en la matière de généralisation possible, les commentaires sur les blogs sont plus « factuels » et attestent d'un réel dialogue entre l'auteur du blog (qui est également son propre « modérateur ») et ses commentateurs, alors que dans un forum l'ensemble des participants et commentateurs (hormis le modérateur) sont « au même niveau » et peuvent intervenir à part égale.

Ainsi la fonction « commentaire » sur les blogs est un hybride idéal entre le web et les forums et listes de diffusion en autorisant simultanément une discussion mono-directionnelle et bi-directionnelle.

26 « Ce qu'on appelle « flux » ou « fil » RSS est un simple fichier texte, comprenant du contenu (en général une succession d'actualités) et des balises délimitant les champs de ce contenu (titres, dates, auteurs). En quelque sorte, c'est un « résumé propre », formaté dans une syntaxe précise, de ce qu'il y a de nouveau sur un site web. Ce fichier est le plus souvent généré automatiquement (...). L'utilisateur qui veut ensuite tirer parti de ce contenu, c'est-à-dire le syndiquer sur son propre site ou simplement consulter le fil d'actualités correspondant, n'a plus qu'à récupérer ce « fichier RSS », accessible par une URL unique, pour l'ouvrir dans un logiciel de lecture (agrégateur ou lecteur RSS). » www.lemondedublog.com/abc

FIGURE 2 – MODÉRATION DES COMMENTAIRES (WORDPRESS)



L'archivage

Un blog se caractérisant par l'affichage ante-chronologique des contenus publiés (le billet le plus récent apparaît en premier), l'ensemble des plateformes proposent depuis l'origine l'archivage automatique de tous ces contenus. Cet archivage peut être mensuel, hebdomadaire ou journalier. Ces trois options peuvent à tout moment être modifiées en un simple clic.

Les catégories

Chaque billet publié peut être rattaché à une catégorie. Certaines plateformes autorisent le rattachement d'un billet à plusieurs catégories différentes. Certaines permettent également d'organiser ces catégories de manière arborescente (à la manière de dossiers et de sous-dossiers).

Quel intérêt ? Les catégories permettent aux lecteurs de naviguer plus aisément à l'intérieur du blog. Elles sont généralement affichées dans l'une des

colonnes latérales de celui-ci. Certains blogs permettent de créer autant de fils RSS que de catégories, ce qui peut s'avérer utile pour mieux cibler la (ou les) thématique(s) du blog que l'on souhaite mettre sous surveillance.

Les permaliens (liens permanents)

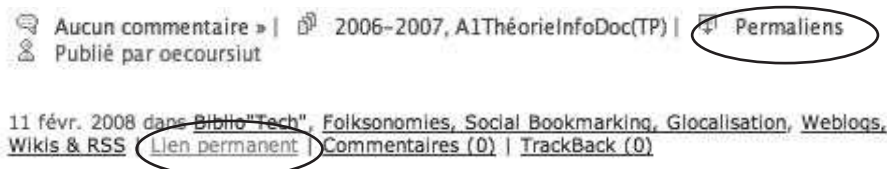
Sur un blog, chaque billet publié se voit affecté une adresse internet (URL) permanente : son permalien.

Quel intérêt ? Cela permet à n'importe quel autre site ou blog de renvoyer précisément vers l'un de vos billets avec la garantie d'éviter les liens brisés suite au déplacement d'une page dans l'arborescence.

Où les trouver ? Les permaliens sont affichés en dessous du billet publié, ou accessibles en cliquant sur la date de publication du billet (*timestamp*).

Comment ça marche ? La procédure est entièrement automatique. Les permaliens générés lors de la publication d'un billet peuvent être de deux sortes : soit ils reprennent les mots du titre du billet (<http://monblog.wordpress.com/le-titre-du-billet.htm>), soit ils affectent un numéro de publication à chaque billet (<http://monblog.wordpress.com/?p=785>).

FIGURE 3 – UN PERMALIEN SOUS WORDPRESS & TYPEPAD



Les rétroliens (*trackbacks*)

De l'anglais *track* (pister, suivre, traquer) et *back* (en arrière, en remontant, à contre-courant), le *trackback* ou « rétrolien » est une technique qui permet à l'auteur d'un blog (A) de placer sur un blog extérieur (B) un lien pointant vers l'un de ses billets (A).

Quel intérêt ? Les rétroliens sont une réelle innovation apportée par les blogs, qui vient pallier l'absence de bidirectionnalité des liens hypertextes traditionnels. Cette fonctionnalité technique distingue les blogs des autres sites web et explique également la très forte densité des liens de la blogosphère.

Comment ça marche ? Concrètement, les rétroliens ne sont possibles qu'entre deux blogs fonctionnant sous des plateformes les autorisant : chaque billet publié est alors doté automatiquement d'une adresse de *trackback* (sous la forme « <http://...> »), adresse différente de celle du billet (son permalien), et

qu'il suffit de copier/coller dans un espace réservé à la fin du billet que l'on est en train de rédiger pour que le lien apparaisse sur le blog cible.

À quoi cela sert-il ? L'intérêt de ces rétroliens est de permettre d'alerter l'auteur d'un billet et ses lecteurs qu'un autre blog, dans un billet précis, traite de la même information, du même sujet ou fait simplement référence à l'autre. Dans le cadre d'une veille ou d'une recherche d'information, il est ainsi possible de découvrir et de repérer de nouveaux blogs, en limitant le bruit informationnel.

Sont-ils fréquents ? Assez répandus et utilisés, les rétroliens ont cependant très vite été détournés pour « spammer ». Dès lors, nombre de plateformes de création et d'hébergement de blogs ont fait machine arrière en imposant une modération *a priori* des rétroliens, mettant ainsi en place un double filtrage (d'abord autoriser les rétroliens sur son blog de manière générale ou billet par billet, puis les valider individuellement pour qu'ils deviennent visibles et actifs).

La *blogroll* ou blogoliste

La *blogroll* est pour un blog l'équivalent d'une page de liens pour un site web. Elle permet de lister les autres blogs que l'auteur souhaite mettre en avant ou sur lesquels il se connecte fréquemment. Généralement affichée dans l'une des colonnes du blog, elle est un élément incontournable de l'identité des blogs. Les *blogrolls* peuvent parfois être organisées à la manière de signets thématiques.

Dans les premiers temps des blogs, et encore aujourd'hui particulièrement pour les blogs d'adolescents, la *blogroll* joue le rôle d'un marquage identitaire sur le mode du « Dis-moi vers qui tu renvoies et je te dirai qui tu es ».

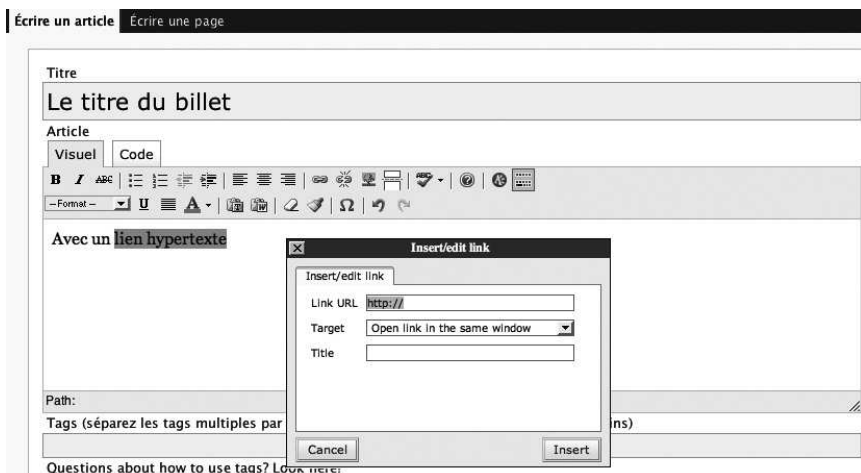
Plus prosaïquement, elle est souvent un moyen de s'inscrire dans une communauté, et de rendre hommage ou crédit à d'autres blogs que l'on tient en haute estime, que l'on juge incontournables, ou vers lesquels on veut renvoyer les lecteurs.

No HTML

La caractéristique technique des blogs est précisément de ne nécessiter *a priori* aucune compétence technique.

Pour le contenu. La publication des billets se fait en mode Wysiwyg, évitant de manipuler le langage HTML. Pour ceux maîtrisant les langages de publication web, il est cependant possible de basculer en mode « code » pour saisir à la main les commandes de formatage et de mise en forme qui ne sont pas disponibles dans l'interface Wysiwyg.

FIGURE 4 – INTERFACE DE PUBLICATION SOUS WORDPRESS



Pour l'habillage et la charte graphique. Les plateformes de blogs proposent le choix entre différents « modèles » (gabarits) à deux ou trois colonnes, ou avec des mises en page plus adaptées aux blogs contenant essentiellement des photos ou des vidéos. Le choix et l'activation de ces gabarits se font sur un simple clic et il est possible d'en modifier très simplement différents éléments (choix de l'image du bandeau de la page d'accueil, taille et type des polices, codes couleurs, etc.).

On peut également accéder et modifier le ou les fichiers CSS²⁷ qui gèrent les gabarits et les modalités d'affichage de la page.

FIGURE 5 – GESTION DE L'HABILLAGE SOUS TYPEPAD

Classique

Ces habillages tirent avantage de tous les modules de colonnes et de l'éditeur de contenu par glisser-déposer.



Mixed Media

Ces formats attirent l'attention sur vos photos et autres médias en les présentant dans une variété de formats attrayants.



Des billets et des pages

Les principales plateformes de blog proposent aujourd’hui deux « espaces » de publication distincts : en plus des billets (affichés de manière ante-chronologique), il est désormais possible de gérer des pages web au sens habituel du terme. Ces pages permettent de s’extraire de la logique d’empilement des billets : en les organisant en arborescence, par exemple, avec indication de pages « mères » et de pages « filles ». Comme sur un billet classique, on peut déposer des commentaires sur ces pages qui disposent d’un permalien, mais on ne peut pas les rattacher à des catégories, ces dernières restant l’apanage des billets.

Attention ! Les fonctionnalités ici décrites reflètent la richesse des plateformes de blogs. Mais elles ne sont pas installées ou permises en totalité par l’ensemble de celles-ci. Pour savoir quelle plateforme autorise quelle(s) fonctionnalité(s), on se reportera aux pages « Créer son blog » de cet ouvrage.

FIGURE 6 – GESTION DES BILLETS ET DES PAGES



Ce qu’il faut retenir

Comme se plaît à le rappeler Joël Ronez, responsable du pôle web de Arte France, les blogs sont « le premier format éditorial né du web ». La principale richesse de ces plateformes éditoriales est de dissocier totalement : la gestion des gabarits (modèles de page, charte graphique), la gestion des fonctionnalités (commentaires, classement et archivage par mots clés et/ou par catégories) et celle des contenus. Il est ainsi possible à tout moment de « remodeler » toute l’apparence ou les fonctionnalités associées au weblog sans que cela nécessite une quelconque remise en forme des contenus publiés.

De la même manière, le déplacement ou le changement de catégorie d’un billet ou d’une page, le changement de nom d’une catégorie, le paramétrage de la modération des commentaires ainsi que l’ensemble des autres paramé-

27 CSS: Cascading Style Sheets: il s’agit d’une syntaxe particulière permettant de gérer des feuilles de style pour l’habillage et la typographie des différents éléments de contenu.

trages possibles s'effectuent sur un simple clic, sans risque d'un effondrement ou d'une perte de données, d'adressage ou de mise en forme.

L'expert comme le novice peuvent ainsi progresser en toute sécurité et avec une réelle latitude d'actions.

Un format de syndication : le RSS

Si la syndication de contenus existait avant même l'explosion des blogs²⁸, le fait que les outils gratuits et grand public permettant de publier des blogs autorisent, pour la plupart d'entre eux, la mise en place automatique d'un ou plusieurs fils RSS va largement contribuer à la promotion de ce format, même si de récentes études²⁹ montrent que les usages sont encore assez peu développés, y compris dans les métiers de l'information-documentation.

◆ Créer son blog

Les solutions « tout en ligne »

Nombre de plateformes (payantes ou gratuites) proposent les deux services de création et d'hébergement en ligne. Votre blog sera alors accessible à une adresse composée comme suit : <http://nomdublog.nomhébergeur.com>. Exemples : <http://volleyclub.wordpress.com> ou <http://volleyclub.typepad.com>.

Quelques plateformes :

- gratuites :
 - Blogger (www.blogger.com) : l'une des plateformes « historiques ». Rachetée par Google, elle est aujourd'hui l'une des plus utilisées mais également l'une des plus limitées en fonctionnalités ;
 - Wordpress (www.wordpress.com) ;
 - Livejournal (www.livejournal.com) ;
- payante : Typepad (www.typepad.com).

Avantages : simplicité et rapidité de mise en place.

Inconvénients : présence de publicité parfois intrusive (sous forme de bandeaux) sur les plateformes appartenant à des sociétés commerciales ; limitation de l'espace disque alloué ; limitation de certaines fonctionnalités (gestion des gabarits, notamment).

²⁸ Pour un aperçu complet du format RSS, consulter notamment : Serge Courrier, *Utiliser les fils RSS et Atom* (voir la référence page 63)

²⁹ Voir à ce sujet le billet de Christophe Deschamps : <http://influx.joueb.com/news/le-rss-peu-utilise-par-le-grand-public-des-chiffres>.

Les plateformes logicielles « clientes »

L'autre possibilité est de dissocier les deux services, en téléchargeant une plateforme de création de blog pour l'installer sur son serveur ou sur celui de son fournisseur d'accès. L'adresse du blog peut alors être plus lisible si l'hébergeur et/ou le FAI permet de créer son propre nom de domaine.

Quelques plateformes :

- Dotclear (www.dotclear.net) ;
- Wordpress (www.wordpress.org) ;
- MovableType (www.movabletype.org/).

Avantages : robustesse et, pour Wordpress.org, des options supplémentaires par rapport à la version « en ligne ».

Inconvénient : nécessite une compétence technique minimale pour pouvoir procéder à l'installation et au paramétrage auprès de son fournisseur d'accès.

Dans tous les cas (outils en ligne ou logiciels clients) et pour bénéficier à la fois d'une meilleure visibilité et d'une meilleure indexation par les moteurs de recherche, il est souhaitable d'acheter son propre nom de domaine et de mettre en place une redirection vers l'adresse à laquelle le blog est hébergé, pour aboutir à des adresses du type : www.volleyclub.fr

Que choisir ?

Le choix d'une plateforme de blog demeure très « personnel ». Il dépend entre autres choses de sa propre compétence technique (suis-je capable de réaliser une installation sur serveur ?) et de la finalité visée (de cette dernière peut dépendre le nombre des options et fonctionnalités nécessaires). D'une manière générale, le choix d'une plateforme cliente offrira toujours davantage d'options à son utilisateur. Mais des services en ligne *open source* comme Wordpress permettent d'obtenir un niveau de qualité presque équivalent.

Il faut donc au final distinguer entre :

- les plateformes communautaires (Skyblogs, Windows Live Spaces, etc.),
- les services gratuits (OverBlog, Vox, Hautetfort, Blogger, Canalblog, Blogspirit³⁰, etc.),
- et les « véritables » logiciels de blogs (Wordpress et Dotclear³¹). C'est vers ces derniers que devra en priorité se tourner le professionnel soucieux d'exploiter toutes les potentialités offertes par les blogs.

³⁰ www.over-blog.com, www.vox.com, www.hautetfort.com, www.blogger.com, www.canalblog.com, www.blogspirit.com,

³¹ www.wordpress.org ou www.wordpress.com, www.dotclear.net

Le site WeblogMatrix³² permet de comparer point par point vingt-trois plateformes de blogs.

FIGURE 7 – COMPARATIF DE QUATRE PLATEFORMES DE BLOGS PROPOSÉ PAR WEBLOGMATRIX

General Features		Blogger	DotClear	Movable Type	Wordpress
URL		www.blogger.com	www.dotclear.net	www.sixapart.com	wordpress.org
Free and Open Source		does not apply	Yes	No	Yes
Multi-User Support		Yes	Yes	Yes	Yes
Anti-Spam		Blogger	DotClear	Movable Type	Wordpress
Blogging Features		Blogger	DotClear	Movable Type	Wordpress
Categorization		None	Single	Multiple	Multiple
Tags		No	Plugin	Yes	Plugin
Archiving		Yes		Yes	Yes
Trackback		No	Send+Receive	Send+Receive	Send+Receive
Link- / Blogroll		Yes	Yes	Plugin	Yes
Protected Posts		No	Plugin	Plugin	Yes
Blog by Email		Yes	Plugin	No	Yes
Post Editing		Blogger	DotClear	Movable Type	Wordpress
WYSIWYG Editor		Yes	Patch	Plugin	Optional
File Upload		Yes	Yes	Yes	Yes
Comment Features		Blogger	DotClear	Movable Type	Wordpress
Feeds for Comments		No	Yes	Plugin	Yes
Extended Features		Blogger	DotClear	Movable Type	Wordpress
Photo Gallery		No	Plugin	Plugin	Plugin
Search Engine		Yes	Yes	Yes	Yes
Custom Pages		No	Plugin	Yes	Yes
Workflow		Blogger	DotClear	Movable Type	Wordpress
Content Approval		No	Yes	Yes	No
Article Publishing		simple	simple	scheduled	simple
Comment Moderation Queue		Yes	Yes	Yes	Yes
Administration		Blogger	DotClear	Movable Type	Wordpress
Themes / Skins		Yes	Yes	Yes	Yes

◆ Quelques blogs

Nous nous essaierons plus tard à une typologie exhaustive des blogs existants mais, pour commencer, attardons-nous sur les quatre grands types de blogs.

Dans la presse

Un temps « débordés » (ou craignant de l'être...) par l'explosion d'une presse communautaire rassemblant les billets rédigés sur différents blogs amateurs³³,

la plupart des sites de presse ont aujourd'hui intégré des blogs dans leur offre éditoriale. Ces blogs sont de deux types différents : les blogs *corporate*, parties intégrantes de la ligne éditoriale du journal, peuvent être le fait d'un chroniqueur ou d'une équipe de chroniqueurs³⁴, les blogs « d'actu » (pour suivre un événement) ou les blogs « dossiers » (pour approfondir une actualité sur la durée) sont systématiquement alimentés par des journalistes ou chroniqueurs. On trouve également une offre réservée aux particuliers³⁵ qui se voient offrir un hébergement pour leur blog, rattaché au site officiel du journal.

À l'université

Les blogs ont aujourd'hui fait leur entrée dans tous les niveaux et auprès de tous les acteurs de l'université. On y trouve des blogs d'enseignants et de chercheurs qui font office de carnets de recherche, des blogs institutionnels (alimentés par le président ou son cabinet, ou bien par un laboratoire de recherche), des blogs d'affichage centrés sur une formation particulière (licence, master, etc.), et enfin des blogs pédagogiques (à usage réservé au dialogue entre un ou plusieurs enseignants et leurs étudiants).

Quelques exemples :

- un blog d'enseignant : Jean Véronis (<http://aixtal.blogspot.com>) ;
- un blog institutionnel : le blog du département de sociologie de l'université Paris-8 (www.univ-paris8.fr/sociologie/) ;
- un réseau de blogs universitaires (enseignants et étudiants) : Université Paris-5 René Descartes (<http://blogs.univ-paris5.fr/>).

En entreprise

On trouve les portails de blogs *corporate* des grands groupes. Parmi eux, des blogs de simples employés ou de responsables à divers postes, des blogs alimentés dans une optique de communication sur un produit, sur un service, ou sur la firme en général³⁶, ou bien encore pour externaliser partiellement ou complètement un processus (comme celui du recrutement³⁷). Le *blogging* d'entreprise comprend donc les « blogs publiés par – ou avec le soutien – d'une entreprise, pour atteindre les objectifs fixés par cette entreprise.³⁸ »

32 www.weblogmatrix.org

33 À l'image du site coréen Ohmynews dont le slogan est « *Every citizen is a reporter* » : www.ohmynews.com/

34 Pierre Assouline (<http://passouline.blog.lemonde.fr/>), Francis Pisani (<http://pisani.blog.lemonde.fr/>).

35 Sur le journal *Libération* : www.liberation.fr/interactif/blogs/, sur le journal *Le Monde* : www.lemonde.fr/web/blogs/0,39-0,48-0,0.html, sur le journal *Le Figaro* : www.lefigaro.fr/blogs/

36 Blog officiel de Google (<http://googleblog.blogspot.com/>), Le blog de l'entreprise General Motors (<http://fastlane.gmblogs.com/>)

37 Voir par exemple Microsoft's Jobsblog : <http://blogs.msdn.com/jobsblog/>

38 www.corporateblogging.info

Certaines entreprises déploient des blogs « dans le cadre » ou « à la place » de leur intranet. D'autres le font de manière publique, dans le cadre d'une stratégie de communication³⁹. C'est ainsi que fin mars 2005 le portail affichant les blogs des employés de Microsoft comptabilisait 1 443 blogs, 34 621 articles (billets) et 120 803 commentaires⁴⁰.

Pour approfondir

Les usages des blogs en entreprise mériteraient à eux seuls un ouvrage. Ils peuvent être déployés en interne ou en externe dans le cadre de différents projets, en support de différents objectifs, en combinant différentes stratégies. En complément des blogs « produits » ou « services externalisés » cités plus haut, on peut également rencontrer des blogs « de communication de crise », des « knowledge blogs » (k-blogs) servant de support à des projets de gestion des connaissances, des blogs dédiés au suivi de la gestion de la relation client (CRM), des blogs « de marque » (pour promouvoir l'image de la marque), etc.

Les blogs en entreprise sont également l'occasion de mettre en place une approche émergente (*bottom-up*) du management par projet (figure 8).

Dans le cadre d'une démarche de veille ou de recherche d'information, et même s'il faut les prendre avec le recul nécessaire⁴¹, les portails de blogs de grandes entreprises (HP ou Microsoft pour reprendre les exemples cités plus haut) constituent un formidable réservoir d'informations auparavant inaccessibles sur la culture de l'entreprise, ses produits, ses méthodes de management, etc.

Les « gourous »

La blogosphère dispose de quelques célébrités, des blogs incontournables du fait de leur audience ou de la médiatisation de leur(s) auteur(s). Loïc Le Meur⁴², Fred Cavazza⁴³ ou encore Laurent Gloaguen⁴⁴ sont quelques-uns de ces leaders d'opinion.

Pour tous les autres, le mieux est encore de consulter les premières places du classement proposé par l'outil Wikio⁴⁵.

39 Voir par exemple le portail des blogs publics de l'entreprise Sun Microsystems : <http://blogs.sun.com/> ou celui de l'entreprise Hewlett-Packard www.hp.com/hpinfo/blogs/index.html

40 Depuis cette date, l'interface du portail <http://blogs.msdn.com> a changé et elle n'affiche plus le nombre de blogs recensés.

41 Les employés, dans le cadre d'un portail de blogs publics hébergés, ne prendront que rarement le risque de dire ouvertement du mal des produits de l'entreprise ou de critiquer ses méthodes de management... Ceux qui s'y sont risqués ont été promptement remerciés.

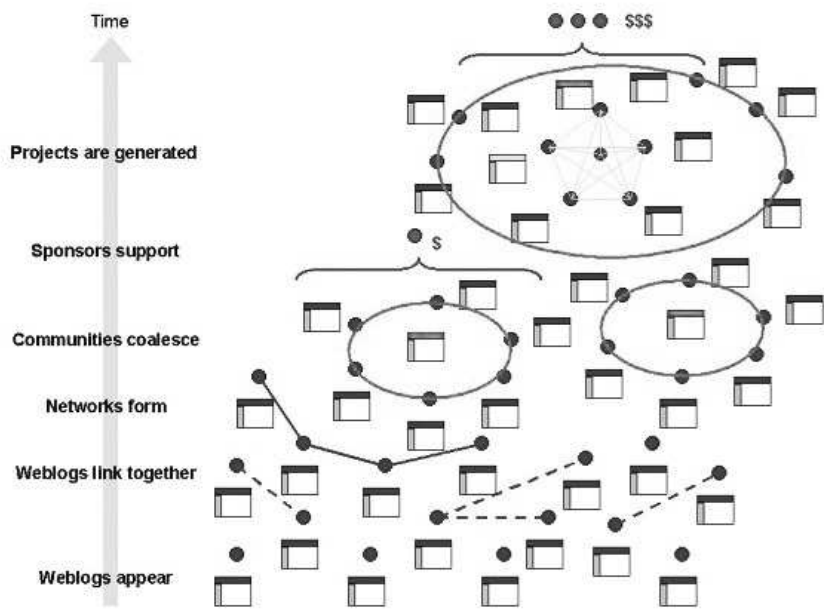
42 www.loiclemeur.com/france/

43 www.fredcavazza.net

44 <http://embruns.net>

45 www.wikio.fr/top/blogs. Voir aussi page 40.

FIGURE 8 – BLOGS... COMMUNAUTÉS... PROJETS⁴⁶



⁴⁶ Source : Martin Roell, *Distributed KM...* (voir la référence page 63).

Chapitre 2

Où trouver des blogs et qu'en faire ?



◆ Combien de blogs ?

Des 23 blogs recensés par un particulier en janvier 1999 aux 70 millions de blogs indexés par le moteur Technorati en avril 2007, la croissance est véritablement impressionnante et leur nombre semble continuer de doubler tous les cinq ou six mois. Ce phénomène n'est pas sans rappeler l'explosion des pages et sites internet aux débuts du réseau. Ce n'est que très récemment que certains analystes ont fait observer un ralentissement dans le rythme de croissance (toujours soutenu) des blogs.

Hormis les précautions d'usage déjà rappelées pour comprendre ces chiffres (tous les blogs ne sont pas « actifs » et la « comptabilité » prend inégalement en compte les différentes plateformes de création et d'hébergement⁴⁷), l'explosion de la blogosphère n'est pas un vain mot. Ainsi, de mars 2003 à juin 2005, un nouveau blog était créé chaque seconde dans le monde, et 13 % étaient mis à jour au moins une fois par semaine, soit près de 900 000 billets par jour⁴⁸ ! Dans l'un des derniers billets de Dave Sifry pour la période d'octobre 2005⁴⁹ à août 2006⁵⁰, cette croissance est confirmée.

La société Digimind, en compilant différentes sources, faisait état dans son dernier Livre blanc⁵¹ d'un chiffre autour de 120 millions de blogs.

À noter. Cette comptabilité des blogs reste essentiellement biaisée par le phénomène des *spam-blogs* (destinés à fausser le référencement des moteurs de recherche) et des *fake-blogs*⁵² (entièrement créés par des programmes pour de la

47 Ainsi, Google indexe « naturellement » mieux les blogs créés sur le service Blogger.com dont il est propriétaire.

48 www.sifry.com/alerts/archives/000332.html

49 www.sifry.com/alerts/archives/000343.html

50 www.sifry.com/alerts/archives/000436.html

51 Christophe Asselin et Véronique Mesguich, *Le web 2.0 pour la veille et la recherche d'information* (voir la référence page 63).

52 www.sifry.com/alerts/archives/000335.html

fraude au clic), sur lesquels nous reviendrons plus tard. Le dernier « état de la blogosphère⁵³ » (avril 2007) indiquait que sur les 120 000 nouveaux blogs créés chaque jour, 3 000 à 7 000 d'entre eux étaient des *fake* ou *spam-blogs*. La détection de ces derniers par le moteur Technorati n'étant pas optimale, on peut raisonnablement penser que ce dernier chiffre était très largement sous-estimé.

Du côté de la Babel linguistique, et toujours selon les chiffres de Technorati d'avril 2007, c'est le japonais qui est la première langue représentée (37 % des blogs), suivi de l'anglais (33 %), du chinois (8 %) et de l'italien (3 %).

◆ Une influence certaine et une légitimité (partiellement) acquise

Dès février 2005, une étude Comscore⁵⁴ révèle que, indépendamment de leur propre courbe de croissance (qui demeure constante), c'est surtout la courbe d'audience des blogs qui devient intéressante. De plus en plus, les blogs sont reconnus parfois à égalité avec les médias « traditionnels ». Le billet « state of the blogosphere » de Dave Sifry indique en septembre 2005 que le pouvoir d'influence (calculé sur le nombre de liens entrants ou *backlinks*⁵⁵) des blogs les mieux positionnés (A-List) par rapport aux médias (comprendre « sites web ») traditionnels se confirme⁵⁶.

Dans un billet du 15 février 2006⁵⁷, Francis Pisani souligne que « les blogs sont plus influents que le *New York Times* ». S'appuyant sur les résultats d'une étude comparant les articles et nouvelles les plus importants de 2007 aux États-Unis, reprenant la liste dressée par les journalistes de l'Associated Press (massacre de l'Université Virginia Tech, crise des hypothèques, guerre en Iraq, prix du pétrole, exportations chinoises), il conclut : « Les blogs gagnent dans toutes les catégories, sauf dans le traitement de la crise des hypothèques et des "supbrimes". »

Dans un tout autres genre, Ebsco – l'une des deux plus grandes agences mondiales de fourniture de bases de données et de périodiques – propose désormais dans son offre la mise à disposition de l'intégralité de blogs « premiums » dans un très large éventail de catégories (environnement, finance, médecine, technologie, etc.). Ceci constitue une nouvelle avancée significative dans la reconnaissance de la valeur ajoutée de la blogosphère, une

53 www.sifry.com/alerts/archives/000493.html

54 www.comscore.com/blogreport/comScoreBlogReport.pdf

55 Les Backlinks désignent les liens entrants vers un site, une page ou un billet de blog. Leur nombre est traditionnellement utilisé pour fournir une indication de popularité ou d'importance du site ou de la page.

56 www.sifry.com/alerts/archives/000336.html

57 <http://pisani.blog.lemonde.fr/2008/02/15/blogs-vs-new-york-times/>

blogosphère en quelque sorte « rééditorialisée » par un éditeur de contenus au sein d'une offre numérique⁵⁸.

Leur nombre, leur influence, leur variété, leur légitimité : ces quatre raisons suffisent à faire de la recherche de blogs une compétence nécessaire et incontournable pour tout veilleur ou professionnel de l'information et de la documentation, et plus globalement pour tout curieux, amateur ou passionné.

◆ Trouver des blogs hier...

Des débuts difficiles

Avec la croissance exponentielle des blogs s'est peu à peu fait jour la nécessité de disposer d'outils permettant de s'y retrouver. Plusieurs soucis techniques spécifiques durent être pris en compte : le premier d'entre eux est que ces blogs furent créés en utilisant différentes plateformes, propriétaires ou non : Blogger, Wordpress, mais aussi les Skyblogs, etc.

Conscients de la volumétrie considérable que les blogs représentaient dans l'ensemble de la galaxie Internet « indexée », les moteurs de recherche les prirent progressivement et naturellement en compte. Mais ils se heurtèrent à une première difficulté : la nature même des blogs étant d'être (ou de pouvoir être) fréquemment mis à jour, et cette fréquence de mise à jour étant par ailleurs l'un des critères importants dans l'ensemble des algorithmes de recherche des différents moteurs, les résultats classiques et la hiérarchie des sites proposés sur une requête – hiérarchie patiemment construite et ajustée – en furent profondément bouleversés.

Les blogs se trouvèrent « surpondérés » du fait de la fréquence de leurs mises à jour d'une part, et du fait, d'autre part, de la densité particulière des liens qui se tissent dans la blogosphère, facilités par la mise en place systématique des *blogrolls* ainsi que par la possibilité des rétroliens. Ces derniers agissent comme autant de *backlinks* (liens entrants), lesquels sont au cœur de l'algorithme de Google et des autres « grands » moteurs. L'un des responsables de celui-ci expliquait début 2005 que « la structure même des blogs (commentaires, billets, catégories, annuaires et classements crawlés régulièrement, phénomène de buzz et de multiplication d'un même lien, fréquence des billets) », favorisait leur bonne position dans les résultats affichés par le moteur.

⁵⁸ Pour les détails de l'annonce d'Ebsco, voir le billet de David Rothman <http://davidrothman.net/2007/03/20/ebsco-to-deliver-blog-content/>

Dès lors il n'était pas concevable sur du long terme de voir des blogs de particuliers spécialistes, par exemple, d'informatique, apparaître classés « devant » des sites institutionnels ou d'entreprise « naturellement » bien classés sur un mot clé générique (« informatique » par exemple) : il en allait de la crédibilité même des résultats proposés par les moteurs, mais également de la viabilité de leur modèle économique.

Syntaxes spécifiques dans les moteurs généralistes

Se trouvant en quelques sorte « pris au piège » de l'explosion de blogs, les moteurs de recherche généralistes se sont d'abord abrités derrière des équations de recherche « avancées » permettant de filtrer les contenus en provenance des blogs au profit de ceux qui provenaient des sites webs « classiques ».

Ainsi, sur Google, il fallait saisir les mots clés suivis de l'expression : « ~blog inurl:archives ». Yahoo! proposait à la même époque une syntaxe identique. Ces deux poids lourds de la recherche s'appuyaient sur la fonctionnalité d'archivage automatique des plateformes de blogs qui associait ainsi systématiquement une zone « archives » à l'URL desdits blogs.

Cette solution ne pouvait être satisfaisante à long terme, mais elle permit un temps de servir de filtre, avant que ne s'impose aux moteurs la nécessité de traiter ces contenus comme une source d'information « spécifique » et de proposer de les interroger *via* un onglet ou une interface dédiée.

Émergence de moteurs spécifiques

Pour pouvoir indexer les nouveaux contenus mis à disposition sur les blogs, de nombreux moteurs « dédiés » ont rapidement vu le jour. Dès 2004, Ari Paparo a maintenu sur son site une liste de ces différents outils⁵⁹, dans laquelle les annuaires cohabitent avec les moteurs (nationaux, internationaux, thématiques ou spécialisés).

La plupart de ces outils avaient une valeur tant que les moteurs majeurs ne permettaient pas d'indexer de manière satisfaisante ces nouveaux contenus. L'indexation des blogs par les grands moteurs sonna le glas de ces « petits » outils qui ne disposaient pas de l'infrastructure nécessaire pour bâtir et mettre à jour des procédures d'indexation en temps réel d'un volume d'information en croissance exponentielle.

⁵⁹ www.aripaparo.com/archive/000632.html

◆ Trouver des blogs aujourd'hui...

Outils de recherche de blogs ou outil de recherche de fils RSS ?

Nombre d'outils de recherche se présentent tantôt comme permettant de trouver des blogs, tantôt des fils RSS. Ils cherchent et indexent en fait la même chose : des sites pourvus de fils RSS, c'est-à-dire principalement et majoritairement des blogs, mais pas exclusivement (les grands sites de presse par exemple disposent de leurs propres fils RSS, et se trouveront donc fréquemment indexés dans des outils de recherche... de blogs).

Si techniquement, pour un moteur de recherche, l'indexation d'un blog (fichier HTML) et celle d'un fil RSS (fichier XML ou RDF) ne sont pas rigoureusement identiques⁶⁰, peu importe en revanche la dénomination moteur de recherche « de blogs » ou « de fils RSS ».

Moteurs généralistes et onglets spécifiques

Google

C'est en septembre 2005 que Google décide de lancer son propre moteur de blogs⁶¹. Les possibilités de recherche avancée sont les mêmes que pour la recherche classique. La seule différence est la possibilité d'effectuer une recherche sur des dates précises de publication (figure 9).

Le nombre de blogs indexés n'est pas communiqué par le moteur, mais sa position quasi hégémonique sur le secteur de la recherche et la possession de la plateforme Blogger en font incontestablement l'un des outils les plus efficaces et les plus conséquents.

Microsoft

Microsoft ne propose pas d'onglet spécifique pour la recherche de blogs. Il est cependant possible d'accéder à une recherche de flux (fils RSS)⁶².

Le moteur généraliste de Microsoft⁶³ permet en revanche depuis janvier 2005 de s'abonner au fil RSS correspondant à la requête saisie. Il suffit pour cela de cliquer sur l'icône affichée dans la barre d'adresse de la page de résultats (figure 10).

60 L'indexation d'un blog peut par exemple permettre de récupérer les informations des balises « Méta », lesquelles ne seront pas nécessairement présentes dans le fil RSS du même blog.

61 <http://blogsearch.google.com> ou <http://blogsearch.google.fr>

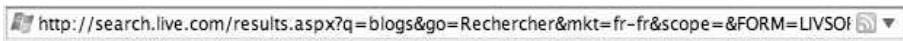
62 À l'adresse : <http://search.live.com/feeds/> (actuellement en version bêta).

63 www.live.com

FIGURE 9 – INTERFACE DE RECHERCHE AVANCÉE DE GOOGLE BLOGSEARCH

The screenshot shows the Google Blog Search advanced search interface. At the top, the Google logo is on the left, and 'Blogs - Recherche avancée' is on the right. Below the logo, it says 'Recherche de Blogs BETA'. The main search area is divided into sections: 'Retrouver des messages' with options like 'tous les mots suivants', 'cette expression exacte', 'au moins un des mots suivants', 'aucun des mots suivants', and 'avec ces mots dans le titre du message'; 'Dans les blogs' with options 'avec ces mots dans le titre du blog' and 'à cette URL'; 'Par auteur' with 'blogs et messages rédigés par'; 'Dates' with 'messages rédigés : à tout moment' and a date range selector; 'Langue' with 'messages rédigés en : Toutes les langues'; and 'SafeSearch' with 'Aucun filtrage' and 'Filtre SafeSearch'.

FIGURE 10 – ABONNEMENT RSS SUR UNE REQUÊTE DANS LIVE.COM



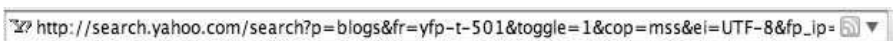
Enfin, ce moteur dispose d'un opérateur de recherche baptisé « feed:votre mot clé ». Mis en place en septembre 2005, il permet de rechercher des fils de syndication parmi les blogs ou les sites web qui en proposent. En saisissant un mot clé après l'opérateur, on récupère la totalité des sites (web ou blogs) possédant un fil RSS et contenant le mot clé saisi dans leur URL.

Yahoo!

Yahoo!⁶⁴ se positionne historiquement du côté des applications sociales et collaboratives, notamment pour se différencier de son concurrent Google. À ce titre, les rachats des sites Del.icio.us et FLickR sont emblématiques. La bascule de Yahoo! vers les techniques de syndication de contenus démarra dès le début de l'année 2005, avec la mise à disposition d'un fil RSS pour certaines rubriques de son annuaire américain. Aujourd'hui, la bascule est complète et c'est l'ensemble des services du portail qui disposent de fils RSS⁶⁵.

En revanche, Yahoo! ne propose pas d'onglet dédié à la recherche de blogs mais, à l'instar de Live.com, il permet de s'abonner au fil RSS d'une requête

FIGURE 11 – ABONNEMENT RSS SUR UNE REQUÊTE DANS YAHOO.COM



⁶⁴ www.yahoo.com

⁶⁵ La page internet <http://developer.yahoo.com/rss/index.html> #biglist résume l'ensemble de ces possibilités.

saisie dans le moteur, en cliquant sur l'icône affichée dans la barre d'adresse de la page de résultats.

L'annuaire (*directory*), service « historique » de Yahoo!, est désormais entièrement accessible sous forme de fils RSS rattachés à chaque catégorie et permettant ainsi d'être tenu au courant des derniers sites ajoutés⁶⁶ (figure 12).

Exalead

Exalead⁶⁷ ne dispose pas d'onglet dédié à la recherche de blogs mais propose depuis peu un service équivalent dans sa « zapette » : la possibilité de choisir les blogs comme « type de site » (figure 13). Il est également possible depuis longtemps d'affiner sa recherche en se limitant aux contenus « syndicables » sous la forme de fils RSS. Cette dernière option permet d'obtenir un spectre de résultats plus large que les simples blogs, nombre de sites du Web 2.0 ou de la presse traditionnelle en ligne étant pourvus de fils RSS.

Ask (blog & flux)

Le moteur de recherche Ask.com, peu utilisé en France mais quatrième moteur « majeur » aux États-Unis, lance le 1^{er} juin 2006 son propre onglet « blogs »⁶⁸ qui donne accès à trois types de résultats de recherche : dans les « billets », dans les « flux » et dans les « actualités » (figure 14).

Il est possible de s'abonner à chaque requête *via* un fil RSS. Le tri des résultats peut s'effectuer par pertinence, par date ou par popularité.

Clusty

Le métamoteur⁶⁹ de recherche Clusty⁷⁰ propose un onglet particulier pour chercher des blogs. Cet onglet permet d'interroger trois moteurs « spécifiques » : Blogdigger, Blogpulse et Technorati.

À noter. Indépendamment des fonctionnalités proposées par les différents moteurs généralistes, il faut noter que les blogs en tant que « source d'information » occupent une place de plus en plus importante de la recherche d'actualités. Les critères d'inclusion d'un blog parmi les sources de « news » indexées sont variables selon les moteurs. Google indiquait ainsi récemment que seuls les blogs comptant plus d'un auteur pouvaient être intégrés dans sa base, une affirmation pourtant démentie par de nombreux exemples.

⁶⁶ <http://dir.yahoo.com/rss/dir/index.php>

⁶⁷ www.exalead.com

⁶⁸ <http://blogsearch.ask.com/>

⁶⁹ Logiciel ou site web interrogeant plusieurs annuaires et moteurs de façon simultanée et en dédoublonnant les réponses.

⁷⁰ Société Vivissimo, <http://blogs.clusty.com/>

FIGURE 12 – FILS RSS DE L'ANNUAIRE ANGLOPHONE YAHOO!

Arts	RSS
Arts > Literature	RSS
Arts > Performing Arts	RSS
Arts > Photography	RSS
Arts > U.S. History	RSS
Business and Economy	RSS
Business and Economy > Business to Business	RSS
Business and Economy > Shopping and Services	RSS
Computers and Internet	RSS
Computers > Video Games	RSS
Computers > Viruses & Worms	RSS
Computers > Weblogs	RSS
Education	RSS

FIGURE 13 – RECHERCHE DE BLOGS ET DE FILS RSS DANS EXALEAD

Préciser la recherche
Termes associés

- [Recherche emploi](#)
- [Moteur de recherche](#)
- [Moteur de recherche google](#)
- [Recherche skyblog](#)
- [Recherche personne](#)

Type de site^{BETA}

- [Blog](#)
- [Forum](#)

Multimedia

- [Audio](#) [Vidéo](#) [RSS](#) [RSS](#)

Langues

- [Français](#)
- [Anglais](#)

Type de fichiers

- [Acrobat \(.pdf\)](#)
- [Text \(.txt\)](#)
- [Word \(.doc\)](#)

[Plus de choix »](#)

FIGURE 14 – RECHERCHE DE BLOGS DANS LE MOTEUR ASK



Moteurs dédiés

Le secteur de la recherche est, dans la blogosphère comme ailleurs, particulièrement concurrentiel. La logique du marché est à la concentration. Ces derniers temps ont vu disparaître, avec l'arrivée sur ce secteur des grands moteurs généralistes, un bon nombre de moteurs dédiés. Après avoir été pour certains incontournables, les moteurs Blogdex, Daypop, Feedster, ou encore

Profusion ont aujourd'hui rejoint le cimetière des moteurs. Il reste heureusement quelques survivants...

Technorati

Sur le secteur de la recherche de blogs, Technorati⁷¹ peut être considéré comme « l'opérateur historique ». Indétrônable jusqu'à l'arrivée de Google BlogSearch, le moteur fait aujourd'hui face à quelques problèmes notamment liés au départ de son fondateur Dave Sifry.

Technorati indique aujourd'hui⁷² indexer 112,8 millions de blogs et près de 250 millions de documents « taggués » (images, vidéos, billets, etc.).

Par défaut, Technorati dispose de quatre onglets permettant de rechercher : des billets (le contenu texte), des blogs (nom du blog et titre des billets), des photos, des vidéos.

Ses atouts

- Recherche des pages pointant vers un blog.
- Recherche par « tags ».
- Recherche de type annuaire (*directory*).
- Possibilité de s'abonner au fil RSS d'une requête.
- Affichage d'un indicateur « d'autorité ». C'est en fait de notoriété qu'il s'agit. L'« autorité » d'un blog sera d'autant plus élevée que le nombre de blogs et de billets le citant sera grand. Cet indicateur permet de filtrer les recherches par « autorité » ou par « fraîcheur » (derniers billets publiés).
- Filtre linguistique (pour chercher, par exemple, uniquement les blogs en français).
- Affichage des tendances (mots clés et requêtes les plus saisis) sous forme de nuages de tags.
- Possibilité de gérer des *watchlists* sur simple enregistrement (gratuit). Il s'agit d'un service d'alerte qui permet d'indiquer des requêtes et d'être tenu informé des derniers billets sur un sujet.
- Dispose d'indicateurs de « classement » rassemblés sous le nom de Top 100. Chaque blog indexé dispose de son « rang » dans le classement (figure 15).

Blogpulse

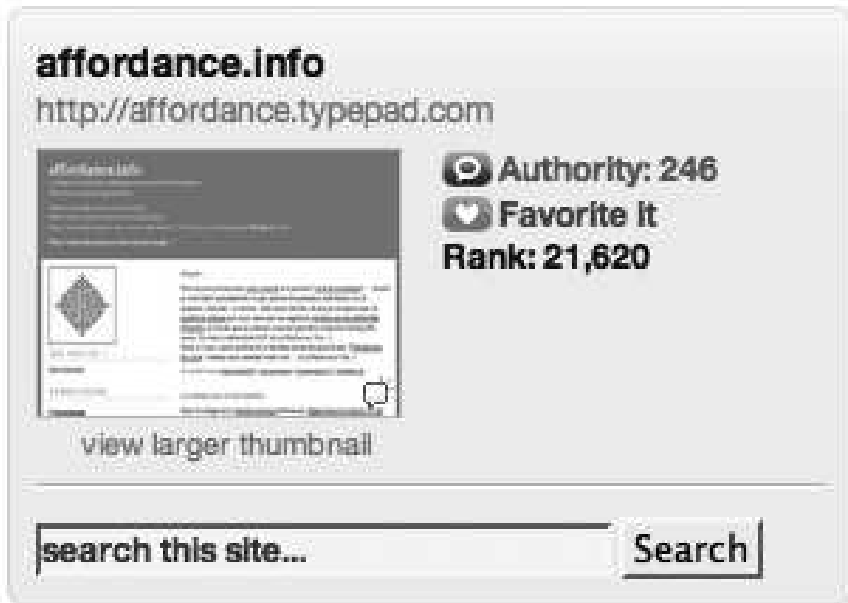
L'outil Blogpulse⁷³ est présenté comme un service de la société Nielsen Buzzmetrics. Ce moteur de recherche se distingue par la richesse des fonctionnalités

⁷¹ www.technorati.com

⁷² <http://technorati.com/about/>

⁷³ www.blogpulse.com

FIGURE 15 – CLASSEMENT ET NOTORIÉTÉ SELON TECHNORATI

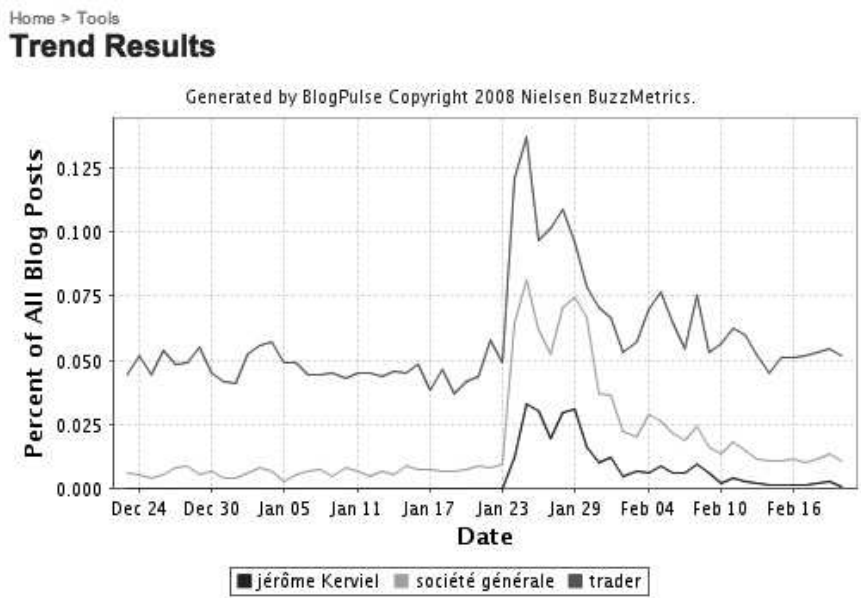


d'analyse qu'il propose. L'outil « Trend Search » permet de suivre, sur une période de six mois, l'évolution du nombre de citations d'un mot clé (ou d'une combinaison de mots clés) dans la blogosphère. L'outil « Conversation Tracker » permet de suivre l'évolution d'une conversation (billets de blogs se répondant ou se citant) à partir d'un mot clé ou d'une adresse de départ. Blogpulse permet également de s'abonner au fil RSS de chaque requête.

Sur les trois courbes représentées en figure 16, on observe très nettement le fonctionnement de la blogosphère comme « chambre d'écho » médiatique à la suite de l'épisode du trader Jérôme Kerviel de la Société générale. Les trois courbes décollent et évoluent de manière similaire à partir du 23 janvier. Un simple clic sur le sommet de la courbe « Jérôme Kerviel » permet d'accéder aux 276 billets qui sont à l'origine de cette poussée.

Le premier lien renvoyé (figure 17) est apparemment tenu par des juristes (son adresse est en effet : <http://lawprofessors.typepad.com>). Un clic sur la fonction « View blog profile » permet d'accéder à une fiche d'identité du blog en question (figure 18). On y apprend ainsi que ce blog, en plus d'être tenu par un professeur de droit (notion d'autorité), bénéficie d'un bon classement – « rank:1000 » (notion de notoriété) et possède en outre une excellente fréquence de publication (notion de régularité).

FIGURE 16 – ANALYSE DU PHÉNOMÈNE « JÉRÔME KERVIEL »



➔ BlogPulse Trend Search

FIGURE 17 – DES COURBES AUX BILLETS DANS BLOGPULSE

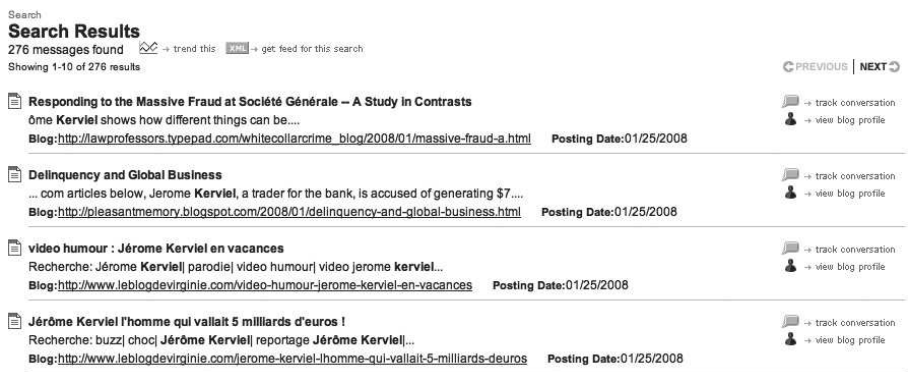
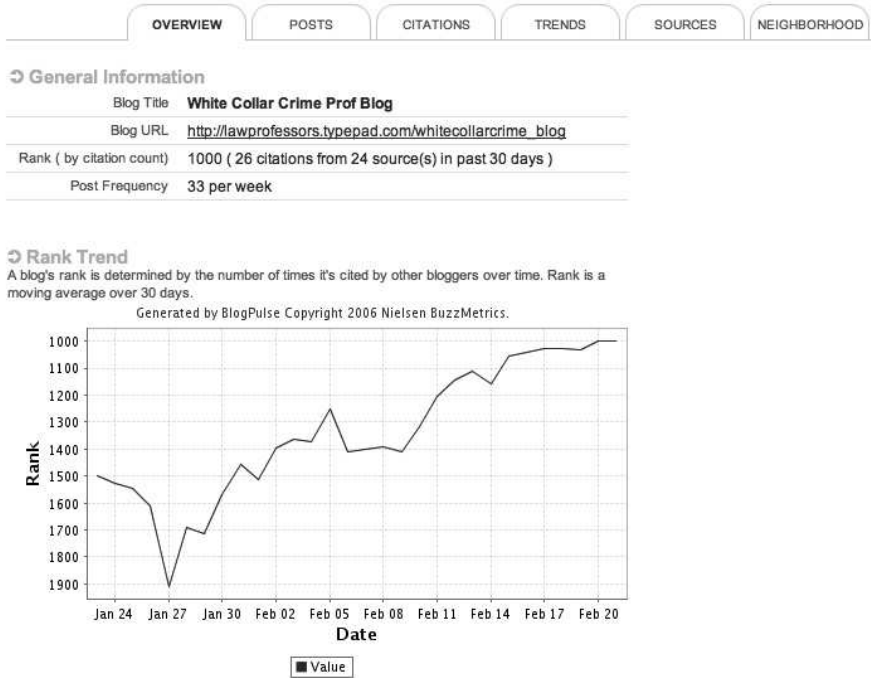


FIGURE 18 – FICHE D'IDENTITÉ D'UN BLOG DANS BLOGPULSE

L'évolution de sa courbe de classement semblant correspondre à l'affaire Kerviel, et compte tenu des critères rappelés ci-dessus (autorité, notoriété, régularité), ce blog constituera donc une excellente source pour analyser et suivre l'affaire Kerviel d'un point de vue juridique. On pourra compléter ce rapide scénario de recherche grâce à l'outil « Conversation Tracker » qui permet de suivre, de « tracer » le chemin des conversations dans la blogosphère à partir d'un terme, d'une adresse de blog ou d'une combinaison des deux.

Dans un autre registre – la rivalité entre les trois grands moteurs de recherche –, une comparaison de leurs courbes de tendance donne le résultat que montre la figure 19. On note ainsi très rapidement que le seul moment où la courbe de Yahoo! et celle de Microsoft rejoignent celle de Google correspond à l'offre de rachat de Yahoo! par Microsoft.

Icerocket

Ce moteur⁷⁴ dispose de cinq onglets permettant de chercher sur le web, dans les blogs et fils RSS, dans les contenus du réseau MySpace, dans les actualités et les images.

⁷⁴ www.icerocket.com

FIGURE 19 – COURBE DES TROIS GRANDS MOTEURS SOUS BLOGPULSE

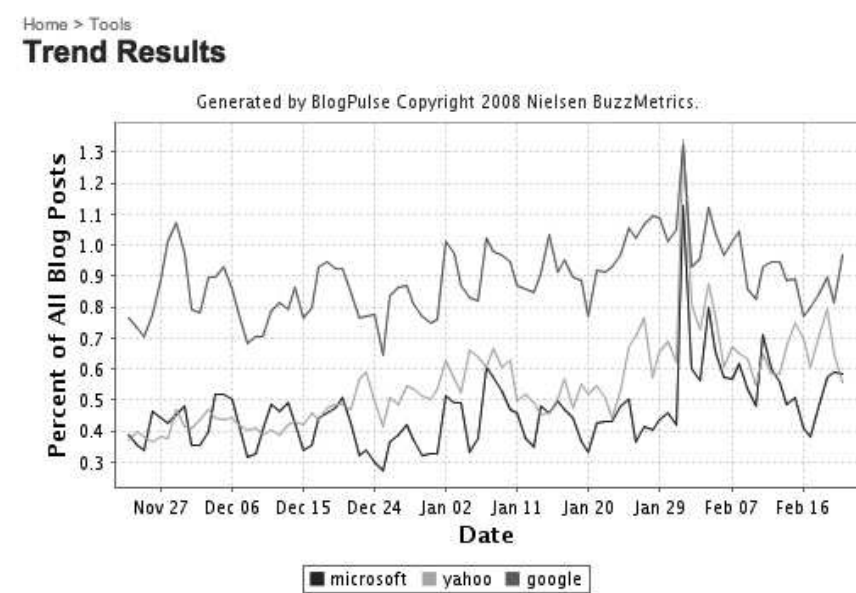
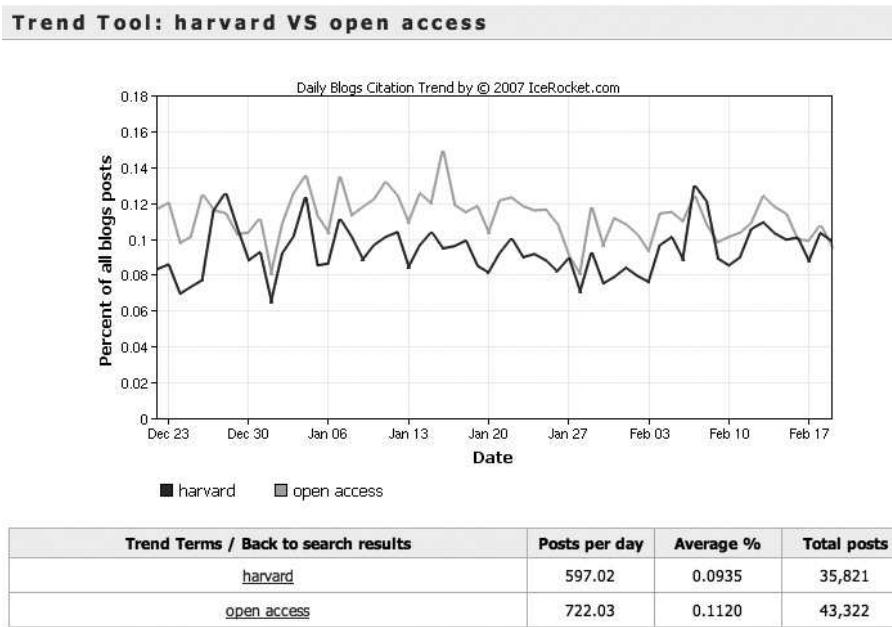


FIGURE 20 – L'OUTIL « TENDANCE » DE ICEROCKET



Ses atouts

- Affichage des mots clés les plus recherchés sous forme de nuage de tags.
- Outils « blog trends » permettant de visualiser des courbes de tendance selon l'occurrence des mots clés saisis. Possibilité de remonter jusqu'aux trois derniers mois (figure 20).
- À la différence de l'outil « Trends » du moteur Blogpulse (cf. ci-dessus), les courbes ne sont pas cliquables et ne fournissent pas d'information contextuelle.

Autres outils

Même si ce n'est pas là l'enjeu du présent ouvrage, il faut rappeler que la recherche d'information et la veille (documentaire ou stratégique) sur les contenus générés par les utilisateurs dépassent de loin les frontières de la simple blogosphère. Nombre de services et d'outils appartenant au Web 2.0 (signets partagés⁷⁵, espaces personnels de partage⁷⁶, hébergement et partage de photos⁷⁷, de vidéos⁷⁸, etc.) proposent un grand nombre de fils RSS généralistes ou spécifiques qui constituent un complément indispensable à toute démarche de veille ou de recherche d'information ciblée.

Bloglines

Bloglines⁷⁹ est un acteur clé du monde des blogs et des flux RSS : ce service permet en effet de chercher des blogs mais également de disposer d'un agrégateur en ligne pour consulter ses fils RSS et les partager avec d'autres utilisateurs. Il permet enfin de créer des blogs hébergés. En février 2005, Bloglines est racheté par le moteur Ask.com.

La particularité de Bloglines est donc de permettre de chercher de manière « communautaire », soit dans la blogosphère en général (en utilisant la technologie du moteur Ask.com), soit dans « ses » blogs ou fils RSS, soit dans les blogs ou fils RSS d'un autre utilisateur (figure 21).

Wikio

Le site Wikio propose un classement mensuel des blogs selon différentes catégories⁸⁰ (figure 22). Ce classement dépend du nombre et de la valeur des liens qui pointent vers un blog sur une période rétroactive de 120 jours. Ce classement thématique peut permettre de délimiter un premier périmètre lors d'une phase de recherche de sources d'informations.

⁷⁵ <http://del.icio.us>

⁷⁶ <http://myweb2.search.yahoo.com/>

⁷⁷ www.flickr.com

⁷⁸ www.dailymotion.com, www.youtube.com

⁷⁹ www.bloglines.com

⁸⁰ www.wikio.fr/blogs/top

FIGURE 21 – INTERFACE DE RECHERCHE AVANCÉE DE BLOGLINES

L'autre point fort de Wikio est qu'il fonctionne comme un agrégateur d'actualités. Disposant de sa propre technologie de recherche, son moteur indexe le contenu de plus de 50 000 blogs. L'outil propose en outre un accès par près de 30 000 catégories réparties en quatorze domaines principaux⁸¹.

Del.icio.us

Le site de signets partagés Del.icio.us (racheté par Yahoo!) compte un nombre important de blogs et de billets de blogs parmi ses signets. En saisissant une adresse de ce type : <http://del.icio.us/tag/mot-clé+blog>, on fait rapidement remonter les blogs traitant de la thématique, du sujet ou du produit que l'on recherche⁸².

◆ Trouver des blogs demain...

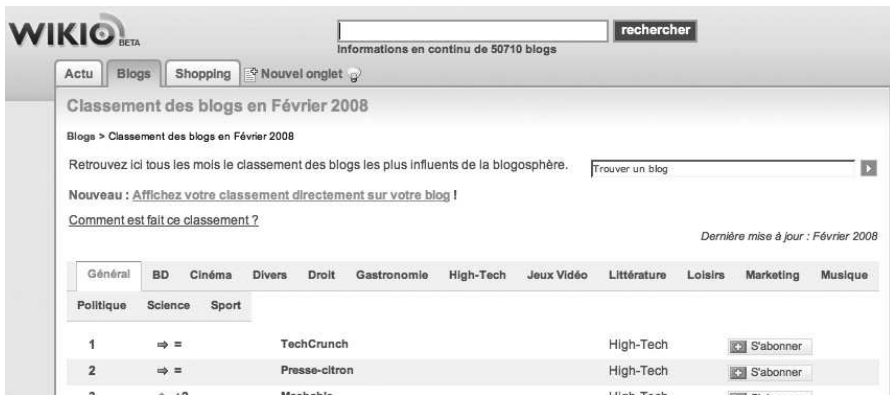
Après avoir été (très) momentanément « déclassés » par les moteurs, après avoir fait l'objet d'un traitement spécifique *via* des opérateurs ou des onglets dédiés, quelle sera la place réservée aux blogs par les moteurs de recherche ? S'il est difficile de le prévoir, deux tendances peuvent cependant être dégagées.

La première est la disparition programmée des moteurs spécifiques au profit des moteurs généralistes.

⁸¹ www.wikio.fr/categories

⁸² Ce scénario de recherche est développé plus bas dans cet ouvrage à l'aide d'un exemple.

FIGURE 22 – CLASSEMENT ET MOTEUR DE RECHERCHE WIKIO



La seconde est celle d'un mixage entre l'ensemble des typologies de documents indexés par les moteurs (pages web, images, *news*, vidéos, blogs), piste présentée par Google sous le nom de « recherche universelle⁸³ ». Du fait de la croissance quantitative de la blogosphère et du fait de sa diversité qualitative, nombre de blogs sont déjà considérés par les moteurs au même titre que des sources journalistiques (et donc accessibles *via* l'onglet « Actualités »).

◆ Blogs : demain... la veille ?

Présentés à leurs débuts comme des réponses possibles au problème de la surcharge informationnelle, les blogs et les technologies de syndication associées doivent aujourd'hui faire face à leur propre « infobésité⁸⁴ ». L'augmentation exponentielle du volume de la blogosphère, l'hétérogénéité qualitative des billets publiés, la diversité des points de vue et des thématiques couvertes, cet ensemble de paramètres fait aujourd'hui des blogs une source d'information parmi d'autres. Pourtant les blogs et la blogosphère disposent de spécificités qui, lorsqu'elles sont bien connues et exploitées, permettent d'optimiser efficacement une simple recherche d'information ou une démarche de veille plus systématique.

83 Voir notamment le dossier que le *Journal du Net* lui consacre (www.journaldunet.com/ebusiness/internet/dossier/070627-recherche-universelle-moteurs-de-recherche/index.shtml), ainsi que ce point de vue plus « critique » : http://affordance.typepad.com/mon_weblog/2007/05/recherche_unive.html

84 Terme québécois désignant une surcharge informationnelle, un trop plein d'information.

Une question de grain : granularité de contenu et d'accès

Le premier intérêt des blogs, dans le cadre d'une démarche de veille, vient de leur structuration même.

Information récente... L'ordre ante-chronologique d'affichage des billets permet d'accéder immédiatement à l'information la plus récemment publiée.

Information archivée... La création de dossiers d'archives mensuels, hebdomadaires ou journaliers est là encore une possibilité offerte « par défaut » par l'ensemble des plateformes de création de blogs. Il devient alors très simple de « s'abstraire » de l'ordre ante-chronologique d'affichage des billets pour naviguer en fonction de la date de publication.

Dans le cadre d'une recherche d'information, l'étendue des archives est un témoin direct de l'ancienneté du blog, sachant qu'il n'est pas aberrant d'établir une corrélation entre « l'ancienneté » du blog et sa pertinence : les blogs alimentés régulièrement et fonctionnant depuis un temps suffisamment long sont généralement le gage d'une information plus pertinente, mieux « sourcée ».

Information catégorisée... La possibilité de rattacher chaque billet publié à une (ou plusieurs selon les plateformes) catégories, offre là encore une possibilité de navigation optimisée pour des blogs multi-thématiques.

Information « cherchable »... La plupart des plateformes permettant de créer son blog proposent également l'installation automatique d'un moteur de recherche⁸⁵. Il est ainsi possible d'accéder, *via* une recherche web standard, à l'ensemble des contenus et des mots clés du blog.

En résumé... Un blog offre donc au chercheur d'information plusieurs entrées croisées :

- recherche « instantanée » : accès aux derniers billets publiés ;
- recherche « temporelle » : accès aux archives ;
- recherche « mots clés » : moteur de recherche ;
- recherche « catégorielle » : consultation et navigation dans les catégories créées.

À cela il faut ajouter la possibilité d'exploiter les liens contenus dans la *blogroll* pour identifier d'autres blogs pertinents sur le sujet, ainsi que les commentaires et les rétroliens qui renvoient généralement vers des blogs appartenant à la même « communauté » ou traitant sensiblement des mêmes sujets.

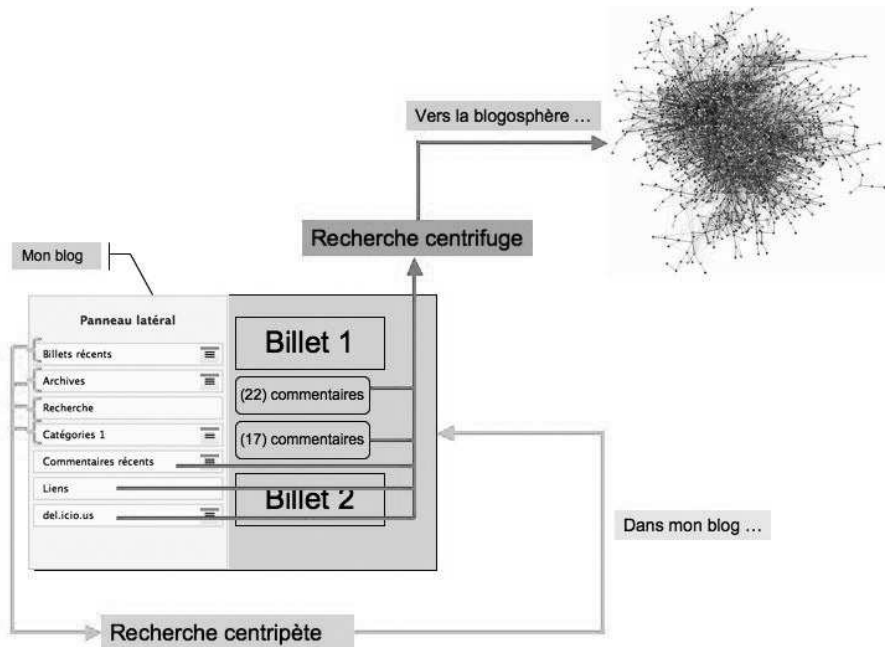
⁸⁵ Pour les quelques plateformes ne proposant pas l'installation d'un moteur de recherche, il est toujours possible, *via* une simple ligne de code, « d'installer » un moteur généraliste dans le cadre de l'offre proposée gratuitement par les principaux moteurs (Google et Yahoo! notamment).

On a ainsi une logique granulaire centripète : recherche « instantanée », « temporelle », « mots clés », « catégorielle » qui permet littéralement de passer le blog au « crible » de notre recherche, et une logique granulaire centrifuge qui permet de remonter le fil des commentaires et des *trackbacks*, vers la blogosphère ou vers d'autres sources.

La principale difficulté restera donc, comme dans toute démarche de recherche d'information, le tri à faire entre le bon grain et l'ivraie, en s'appuyant pour cela sur l'identification systématique des « sources » et des auteurs présents derrière le blog⁸⁶, et en croisant les informations obtenues avec d'autres sources. À ce titre, et comme rappelé ci-dessus, les différents éléments de granularité d'accès peuvent constituer une aide précieuse dans l'identification et la labellisation des sources.

Enfin, pour une veille efficace sur la blogosphère (et bien au-delà), il faut souligner le rôle essentiel que jouent les différents formats et technologies de syndication de contenus, et ce dans les phases de repérage, d'identification, d'analyse et d'exploitation des informations⁸⁷.

FIGURE 23 – « UN » BLOG POUR « DES » RECHERCHES



86 Chaque blog propose par défaut une page « À propos » qui permet aux auteurs qui le souhaitent de s'identifier et de livrer leurs motivations et/ou la ligne éditoriale de leur publication.

87 Voir : Serge Courrier, *Utiliser les fils RSS et Atom* (référence page 63).

Blogs et recherche d'informations : en bref...

La fréquentation assidue de la blogosphère peut s'avérer rapidement étouffante tant les sources d'information y sont abondantes et renouvelées, et ce sur une très riche variété de thèmes et de domaines. Les quelques conseils ci-contre⁸⁸ permettront d'atténuer sensiblement la sensation d'infobésité.

À noter. Si les blogs peuvent occuper une place centrale dans une recherche d'information en tant que « source », ils peuvent également être l'un des éléments à verser au dossier de veille en tant que document de référence. Ainsi, leur structure et leurs fonctionnalités en font des supports très efficaces pour la constitution d'un dossier documentaire⁸⁹.

L'effet *pundit* pour la veille et la recherche d'information

Ou comment le « chercheur » devint « trouveur ».

À condition de pratiquer cet exercice depuis un certain temps et avec régularité, l'auteur d'un blog peut bénéficier de ce que la littérature qualifie d'« effet *pundit* », c'est-à-dire une expertise affirmée qui le place au cœur d'un réseau (ou collège invisible) et d'une communauté, lesquels deviennent ses principales sources d'information et lui permettent d'être alerté sur les dernières nouveautés de son secteur en mode *push*, c'est-à-dire sans avoir à effectuer lui-même une démarche proactive de recherche. En d'autres termes, sans en « chercher » réellement, il « trouve » de l'information avec un gain quantitatif et qualitatif indéniable.

Les *pundits* sont des experts (parfois auto-proclamés) qui drainent vers eux et sur un domaine de compétence plus ou moins étroit les avis et commentaires d'autres experts. Les commentaires sont à ce titre la zone prioritaire dans laquelle collecter ces remontées d'informations. Par ailleurs, les petits *widgets*, utilitaires permettant de visualiser en temps réel ou en différé l'origine des visiteurs sur un blog⁹⁰, permettent à leur tour de repérer d'autres blogs qui sont eux-mêmes des sources d'information incontournables ou complémentaires sur les mêmes sujets.

Nous sommes donc ici en présence d'une communauté d'intérêt générée et non imposée par une quelconque hiérarchie ou administration, ce qui fait à

88 Dont certains sont inspirés du billet de Bill Burnham : http://billburnham.blogs.com/burnhamsbeat/2005/11/feed_overload_s.html

89 Pour un exemple, voir « Comment utiliser les blogues pour la diffusion de dossiers documentaires » http://dossierdoc.typepad.com/dossierdoc/2004/12/comment_utilise.htm

90 Exemples de services: Referer.org (www.referer.org), Mybloglog (www.mybloglog.com/), Criteo widget (<http://widget.criteo.com/>)...

DES CONSEILS POUR LA RECHERCHE DE BLOGS

PENDANT LA RECHERCHE	
Exploiter les moteurs de recherche	Exploiter les opérateurs spécifiques des moteurs en proposant.
	Rechercher les mots clés dans le titre et/ou le contenu des billets
	Croiser les moteurs généralistes et les moteurs dédiés
	Combiner avec une recherche dans des outils « sociaux » de type Del.icio.us
Exploiter la structure des blogs et de la blogosphère	Recherche « instantanée » : accès aux derniers billets publiés
	Recherche « temporelle » : accès aux archives
	Recherche « mots clés » : moteur de recherche
	Recherche « catégorielle » : consultation et navigation dans les catégories créées
	Suivi des commentaires
	Exploration de la <i>blogroll</i>
Préférer les blogs éditorialisés aux « revues de liens »	Gérer le ratio entre les blogs d'éditeurs (publishers) et ceux de « distributeurs » (<i>linkers</i>) : les premiers publient (assez) peu mais des billets relativement longs et « originaux ». Les seconds publient énormément de billets dans lesquels ils se contentent de reprendre des billets glanés ailleurs assortis de leurs propres commentaires. Une recherche d'information ou une veille documentaire ou stratégique se focalisera davantage sur les blogs de <i>publishers</i> afin de rassembler les principales informations et d'identifier les principaux acteurs sur un sujet ou une thématique donnée. En revanche, une veille image ou de positionnement marketing s'attachera à débusquer le maximum des <i>linkers</i> afin d'évaluer de manière quantitative la viralité d'une campagne publicitaire ou d'un produit.
AVANT ET APRÈS LA RECHERCHE	
S'organiser par « dossiers »	Regrouper et rediriger ses blogs dans des « dossiers » auxquels peuvent être aisément mis à jour et où il est facile de se repérer de manière thématique ou en fonction de la tâche.
Viser le niveau « méta »	S'abonner à des méta-blogs : certains blogs verticaux proposent en un endroit unique le contenu de plusieurs blogs (ou de certaines rubriques d'autres blogs) sur une thématique ou un secteur donné. Idem pour les moteurs de recherche à partir desquels il est possible de créer des fils de syndication sur tel tag ou tel mot clé.
Purger et réorganiser fréquemment les blogs surveillés	Restructurer ses dossiers et sous-dossiers, remplacer les blogs peu consultés par des méta-blogs sur les mêmes thématiques.

la fois sa force et sa pertinence. Car ces blogs découverts soit *via* les commentaires de leurs auteurs, soit par l'affichage de la provenance des visites, sont très souvent ceux dans lesquels l'auteur du blog « cible » puise bon nombre d'informations ou d'idées pour ses propres billets et ses propres recherches. Ils sont également ceux à partir desquels il est possible de découvrir (*via* leurs propres listes de liens ou *blogrolls*) les autres blogs restant à visiter dans une démarche cette fois plus proactive.

Ces *pundits* bénéficient donc d'un cercle vertueux qui renforce leur autorité en même temps qu'il leur permet de découvrir de nouvelles sources et d'alimenter leur propre réflexion.

Ce phénomène s'apparente à l'effet de seuil de la sérendipité⁹¹ : celui à partir duquel le blogueur *pundit* dispose de suffisamment d'informations et de sources permettant de renvoyer vers d'autres dont la majorité s'avérera pertinente et dans lesquelles le bruit informationnel sera quantité négligeable.

Une technologie incitative qui rend bavard : l'affirmation des expertises

Du fait de leur facilité de mise en place, d'usage et d'alimentation, les weblogs sont un facteur d'incitation qui met à disposition des veilleurs de nouvelles sources, pour la seule raison que de nouveaux individus se mettent à publier des informations qui réclamaient auparavant de passer des filtres techniques démotivants (connaissance du code HTML, gestion d'un espace d'hébergement, gabarit du site et de la page, etc.).

Quelle que soit leur provenance, et à tous les échelons hiérarchiques de l'entreprise, de l'université, de la presse, ces « nouveaux » acteurs de la diffusion d'information publient et communiquent avec, la plupart du temps, une liberté de ton que rend possible l'absence de processus de validation ou de chaîne éditoriale.

Ainsi journalistes, universitaires, mais aussi PDG, CEO et autres experts sectoriels deviennent de nouvelles sources disponibles pour les veilleurs. Des gens jusqu'ici détenteurs d'un savoir opérationnel ont la possibilité de le rendre disponible à leur communauté ou au plus grand nombre.

91 Découverte par chance ou par sagacité d'informations que l'on ne cherchait pas. Pour les applications de la sérendipité à la recherche d'information, voir Olivier Ertzscheid *et al.*, *Perspectives documentaires sur les moteurs de recherche : entre sérendipité et logiques marchandes*, http://archivesic.ccsd.cnrs.fr/sic_00172169/fr/

Un nouvel écosystème de veille ?

Nombre d'entreprises, de cellules de veille, de documentalistes et de particuliers adeptes des technologies RSS directement liées aux blogs, reconnaissent avoir observé un gain qualitatif et quantitatif significatif pour leurs activités de veille ou de recherche d'information. Le gain de temps que permettent les agrégateurs est incontestable : le *monitoring* de plusieurs dizaines de blogs et autres sources d'information disposant d'un fil RSS ne nécessite que quelques secondes. Qualitativement, et pour peu que le travail de sélection des sources ait été fait de manière intelligente, le gain est également très important.

Le principal intérêt des blogs et du RSS réside dans leur contribution à un changement de paradigme : nous sommes en train de vivre et d'expérimenter le passage d'une logique de « puits » (celui qui a l'information a le pouvoir) à une logique de « nœud » (celui qui distribue l'information a le pouvoir). Le veilleur qui « capte » les relais, les nœuds les plus adaptés à son sujet dispose ainsi d'un temps d'avance.

Il ne faut pas pour autant idéaliser les pratiques de veille liées aux blogs et au RSS. En septembre 2005, je publiais un article⁹² indiquant toutes les potentialités des blogs et du RSS pour la veille. Plus de deux ans plus tard, la dernière étude de Digimind⁹³ montre à quel point l'adoption de ces pratiques peut encore être améliorée : « Plus de 50 % des cellules de veille ne surveillent pas ou peu les forums, blogs et autres réseaux sociaux. »

92 Olivier Ertzscheid, « Weblogs : demain... la veille » (voir la référence page 63).

93 *Baromètre 2007 des pratiques de veille des grandes entreprises françaises*, www.digimind.fr/publications/etudes-de-marche/341-barometre-2007-des-pratiques-de-veille-des-grandes-entreprises-francaises.htm

Chapitre 3

Quelle typologie pour quels contenus ?



◆ Quelle typologie ?

Si la nature des blogs reste largement protéiforme, il est cependant possible d'en dresser une typologie, dans une démarche de recherche d'information et de meilleur ciblage des sources à consulter.

Journaux personnels / Carnets de notes / Méta-blogs

Une première typologie permet de distinguer entre trois types de contenus différents⁹⁴.

◆ On trouve d'abord les **journaux personnels**, égotistes, simples « consignes » d'états d'âme quotidiens. Dans cette catégorie trouvent place les très nombreux Skyblogs. C'est le cas d'une très large majorité des blogs créés, et ce indépendamment de la plateforme de création et d'hébergement choisie.

Si l'immense majorité des blogs composant cette sphère est celle des journaux intimes, tenus sur l'internet de manière individuelle, on y trouve également les blogs familiaux qui permettent à des familles géographiquement dispersées de « rester en contact » en diffusant, par exemple, des photos ou des vidéos. Un « sous-genre » de ces blogs familiaux a connu un essor particulier aux États-Unis : il s'agit du phénomène des *babyblogs*, dans lesquels on poste des photos du dernier-né. Il existe enfin des blogs consacrés à un loisir (cuisine, bricolage, philatélie, etc.).

◆ Viennent ensuite les **carnets de notes** (*notebooks*) qui proposent un contenu sous forme de points de vue, d'analyses et de synthèses (*focused essays*) sur

94 Cette première typologie est inspirée des travaux de R. Blood, *The Weblog Handbook: Practical Advice on Creating and Maintaining your Blog*, Cambridge, MA, Perseus Publishing. 2002.

des sujets de société ou d'actualité. Comme pour les journaux personnels, le spectre est ici très large et l'on peut aussi bien y compter les blogs traitant de recettes de cuisine que ceux de scientifiques parlant de leurs recherches, ou de journalistes s'y livrant à une activité de chronique.

Il faut y distinguer deux sous-ensembles :

- ceux qui sont tenus par des individus isolés : ils peuvent ne pas avoir de rattachement institutionnel ou organisationnel lisible, ou bien être affiliés, par exemple, au site d'un grand quotidien ou d'une entreprise. Nombre de grands quotidiens offrent ainsi un blog à certains de leurs journalistes ou de leurs éditorialistes (Assouline, Pisani, Schneidermann, etc.) ;
- ceux qui sont alimentés collectivement dans une logique de soutien, de valorisation ou d'information sur une association ou une action humanitaire (Secours populaire, réseau Éducation sans frontières, etc.).

Ce qui permet de les distinguer des blogs de la première et de la troisième catégories, c'est la notion d'« éditorialisation » et le focus « thématique ». On y perçoit un « ton », un « style » qui peut porter sur le fond (les sujets traités) ou sur la forme (la manière dont l'auteur s'exprime).

◆ Viennent enfin les **méta-blogs** ou **blogs-filtres** (*filters*) qui se subdivisent en deux sous-ensembles :

- ceux qui, au hasard des découvertes de leur auteur, se contentent de picorer et de reprendre des contenus en provenance d'autres blogs, soit sous forme de citations ou de liens hypertextes, soit en reprenant l'intégralité du billet (il peut alors s'agir de cas avérés de plagiat si le contenu des blogs « sources » n'est pas distribué sous une licence adaptée⁹⁵) ;
- les « vrais » méta-blogs qui, sur un sujet donné, rassemblent de manière systématique et ordonnée tout ou partie des contenus d'autres blogs. Ces blogs sont un peu l'équivalent des moteurs verticaux ou thématiques qui ne donnent accès qu'à un seul type de ressource sur une seule thématique. Si, quantitativement, ces derniers restent largement minoritaires au regard des deux catégories précédentes, leur valeur ajoutée qualitative pour une recherche d'information est en revanche essentielle. Le méta-blog Planète Web Sémantique⁹⁶ en est un bon exemple.

⁹⁵ Le contenu de beaucoup de blogs est accessible sous une licence Creative Commons permettant à n'importe qui d'en reprendre le contenu sous certaines conditions (par exemple en citant l'auteur, ou en dehors de tout cadre commercial).

⁹⁶ <http://planete.websemantique.org>. Comme on peut le lire sur l'écran d'accueil, ce blog « rassemble les principaux carnets web d'auteurs francophones traitant du Web sémantique. Ces blogs sont souvent bilingues français/anglais et éclectiques. Cette page rassemble l'ensemble de leurs billets et vous permet de sélectionner ceux que vous voulez voir. » Ce blog propose également des signets del.icio.us indexés avec des mots clés liés au Web sémantique.

Mais les authentiques méta-blogs francophones sont encore relativement rares. En revanche, la blogosphère est, à l'intérieur de ses propres limites, une formidable chambre d'écho : nombre de blogueurs se contentent de reprendre des informations publiées par d'autres, plus influents, jouissant de davantage d'autorité ou tout simplement mieux informés, en y ajoutant une « touche personnelle » souvent de peu d'intérêt. Les méta-blogs permettent de mieux cerner cette propagation virale de l'information, intrinsèque à la blogosphère, en évitant de systématiquement s'y perdre.

Réflexions / Analyses / Signalements

Une autre manière de « lire » ou d'interpréter la typologie ci-dessus est de considérer que les journaux intimes proposent des *réflexions*, là où les carnets de notes s'efforcent de mettre en avant des *analyses*, les méta-blogs se contentant de *signalements*.

Blogs personnels ou collectifs, thématiques ou intimes

Une autre grille d'analyse intéressante pour comprendre et positionner ces sources d'information est de les considérer selon deux axes. Sur le premier axe, on place le contenu des blogs sur une échelle qui va du personnel au thématique (*topical*). Sur le second axe, on prend en compte le nombre d'auteurs sur une échelle qui va de l'auteur unique (cas le plus fréquent) aux blogs collectifs multi-auteurs.

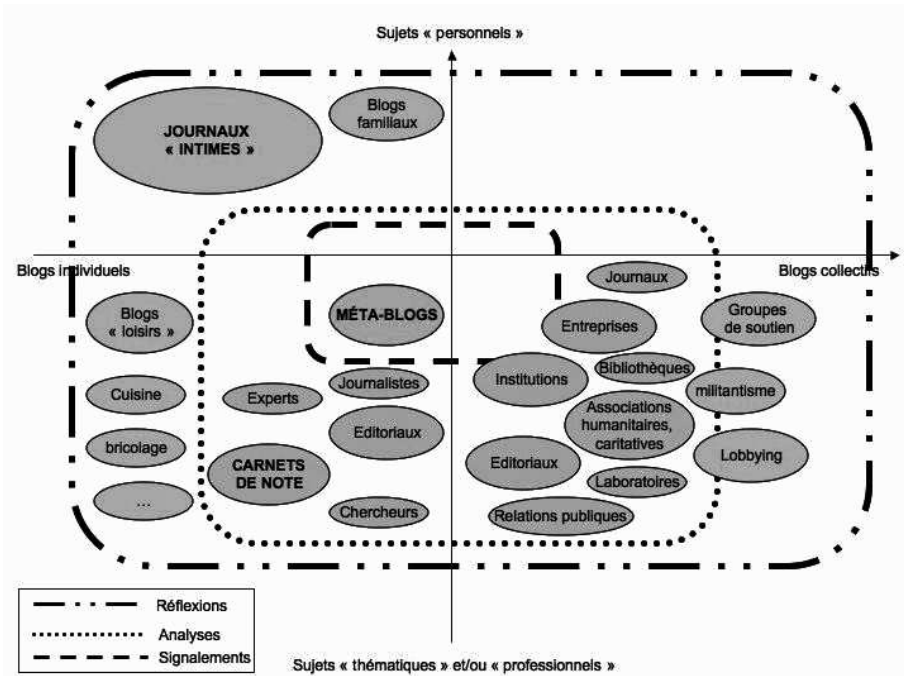
Vue synthétique

La figure 24 propose de rassembler l'ensemble des approches présentées jusqu'ici.

Une réalité diversifiée

De ne pas entrer « simplement » dans l'une des catégories présentées ci-dessus fait toute la richesse et la difficulté des blogs. Dans les faits, de nombreux blogs sont une hybridation de ces différentes catégories. On rencontre fréquemment des blogs à dominante carnet de notes mais qui, selon les périodes ou les sujets, fonctionnent davantage à la manière de méta-blogs (se contentant de signaler d'autres ressources) tout en s'autorisant de fréquentes incursions du côté du journal intime (en relayant les états d'âme de leur auteur).

FIGURE 24 – UNE TYPOLOGIE DE LA BLOGOSPHERE



Dans le cadre d'une recherche d'information, la question est donc bien de déterminer une majeure et une mineure, afin de permettre de jauger l'intérêt de tel ou tel blog en fonction de sa propre stratégie de recherche.

Enfin il faut noter que si actuellement cette hybridation, ce « mélange des genres » reste opérant, on observe cependant – notamment au travers des sites de classement de blogs – la stabilisation d'un certain nombre de thématiques. Ainsi les blogs « scientifiques » apparaissent comme une dominante assumée, au même titre que les blogs « de cuisine », les « BDblogs » ou les blogs « high-tech ».

Si l'on s'en tient par exemple à l'observation des blogs de la sphère politique, on trouvera aussi bien des blogs alimentés par des élus qui y expriment leurs états d'âme, y tiennent leur carnet de campagne ou reviennent sur la politique qu'ils mènent, que des blogs tenus par l'équipe politique d'une personnalité de premier plan ou encore des blogs créés et alimentés par des groupes de soutien à tel parti ou à tel candidat.

◆ Petite sociologie du blogueur

Au travers des exemples précédents, on aura compris toute la difficulté d'établir un profil-type du blogueur. Une étude de l'institut Pew Internet datant de janvier 2005 permettait de faire remonter différents critères de caractérisation. Le blogueur-type serait : plutôt un homme (57 % des sondés); plutôt jeune (48 % des sondés ont moins de trente ans); avec une bonne connexion Internet (70 % des sondés disposent du haut débit chez eux); vétéran de l'Internet (82 % des sondés surfent sur le Web depuis plus de six ans); avec une situation financière plutôt aisée et/ou un niveau d'éducation supérieur.

La plupart des autres études font état de résultats similaires⁹⁷, le seul paramètre revu significativement à la baisse étant celui de la masculinité (le nombre de blogs féminins est en effet de plus en plus important).

◆ Le profil-type du chercheur de blogs

Une étude de 2006⁹⁸ s'est intéressée aux requêtes déposées sur des moteurs de recherche de blogs (moteurs dédiés) en les comparant à celles des moteurs généralistes. Cette étude permet non seulement de mettre à jour un profil-type du chercheur de blogs, mais également de faire émerger les spécificités de la blogosphère en tant que « sous-ensemble » du Web.

Tout d'abord, en ce qui concerne les requêtes, l'étude isole deux catégories : les requêtes contextuelles (mots clés contextualisés : « 11 septembre », par exemple) et les requêtes conceptuelles (mots clés classiques : « attentats », « terrorisme »). Cette première distinction n'est pas spécifique aux blogs et se retrouve également dans l'interrogation de sources « web ». En revanche, les auteurs signalent une tendance très nette des requêtes portant sur des événements courants ou récents. Chercher « live » plutôt que chercher « wide »⁹⁹.

Du côté des comportements de recherche, comme lorsqu'ils utilisent des moteurs généralistes, les usagers ne prêtent attention qu'aux tout premiers résultats retournés et ne renouvellent que très peu leurs requêtes lors d'une même session : si les résultats restent insatisfaisants, ils passent à autre chose. En conclusion, le chercheur de blog apparaît comme quelqu'un qui est « beaucoup plus tourné vers les actualités (politiques en particulier), les loisirs et la technologie que le chercheur web lambda. »

97 Dont celle menée en 2005 par une chercheuse du MIT-MediaLab (<http://alumni.media.mit.edu/~fviegas/survey/blog/results.htm>)

98 Gilad Mishne, Maarten de Rijke, *A Study of Blog Search*, <http://staff.science.uva.nl/~gilad/pubs/ecir06-blogsearch.pdf>

99 En référence au bandeau d'accroche du moteur Technorati qui proclamait : « Search the world live web. »

C'est donc *in fine* la nature première de l'information qui circule dans les blogs que permet de retrouver cette étude, par un simple effet miroir. Les thématiques effectivement les plus couvertes par les blogs sont bien celle des loisirs, de l'actualité politique et de la « high-tech ».

◆ Approfondir...

Le brevet de Google

Si en tant que « contenus » les blogs ne posent aux moteurs aucun problème particulier d'indexation, il n'en est pas de même pour les logiques de classement et d'affichage. Les spécificités des blogs tiennent en effet principalement à la fréquence de leurs mises à jour, critère qui, avant l'explosion de la blogosphère, figurait en bonne place dans l'algorithmie des moteurs pour construire le classement de leurs résultats.

Attardons-nous un instant sur le brevet déposé par Google le 13 septembre 2005, publié le 15 mars 2007 et intitulé *Ranking Blog Documents*¹⁰⁰. Les critères suivants sont retenus : « popularité du blog (...), présence du blog dans diverses *blogrolls*, présence du blog dans des *blogrolls* de blogs eux-mêmes populaires et de qualité (...), liens externes pointant vers le blog en provenance d'autres sources, *pagerank* du blog, lectorat RSS. »

Au final, des critères assez « classiques » mais qui attestent une nouvelle fois de l'importance de l'étude des blogolistes (*blogrolls*) dans une démarche de recherche d'information ou de veille afin d'identifier les experts ou les blogs incontournables du domaine sur lequel porte la recherche.

Le cimetière des blogs

Avec l'engouement suscité par les blogs, avec la facilité et l'instantanéité de leur mise en ligne, nombre de personnes créent leur blog « juste pour voir ». Elles sont ensuite une majorité à ne l'alimenter que très peu ou pas du tout. Ainsi un très grand nombre de Skyblogs ne contiennent en tout et pour tout qu'un seul billet (celui de leur mise en ligne). Dans un autre registre, il faut citer les nombreux blogs d'étudiants et/ou d'enseignants créés dans le cadre de cours ou de travaux pratiques et qui ne sont alimentés que pendant ce temps pour ensuite être abandonnés. Or, dans un cas comme dans l'autre, très peu de personnes font l'effort de supprimer un blog qu'elles n'alimentent pourtant plus.

¹⁰⁰ Ce brevet est consultable sur la base UPSTO à cette adresse : <http://tinyurl.com/35etd4>

Le problème est que ce cimetière des blogs entre pour une bonne part dans l'inflation des chiffres de la blogosphère. Comme le rappelait Francis Pisani dans l'un des billets de son blog, « quand on se penche sur les statistiques de population, on ne compte pas les morts. Pourquoi le faire pour les blogs? » Tout cela constitue une source de bruit informationnel considérable pour les outils de recherche... et pour le chercheur d'information.

Blogs et splogs : un écosystème fragile

À l'image des courriels et de leurs ennemis jurés, les pourriels (courriers indésirables), les blogs disposent de leurs propres logiques de spam. Le spam peut s'infiltrer dans les espaces de dialogue et d'interactivité laissés ouverts (commentaires, *trackbacks*) ou bien se manifester directement par le création de splogs (*spam-blogs*) ou fblogs (*fake-blogs*, « faux-blogs »).¹⁰¹

Les splogs sont principalement destinés à fausser le référencement des moteurs de recherche, au profit de celui qui les met en place ou à tout le moins au détriment de son (ou ses) concurrent(s). Les fblogs sont, eux, principalement destinés à générer de la fraude au clic.

Spam de commentaire et de *trackback*

Les programmes qui parcourent le Net à la recherche d'adresses mail pour effectuer ensuite du spam sont à peu de choses près les mêmes qui viennent automatiquement déposer en commentaire (ou en rétrolien) des liens pointant vers des sites aux contenus indésirables. Nombre de blogueurs se sont un matin réveillés en constatant que leur compteur de commentaires avait littéralement « explosé », enregistrant en quelques heures le dépôt de plusieurs milliers d'entre eux. Chaque plateforme de blog a dans un premier temps proposé à ses utilisateurs d'installer un *patch* ou *plug-in* permettant un niveau de filtrage plus efficace. Mais, indépendamment des désagréments pour l'utilisateur et l'auteur du blog, les moteurs de recherche furent les premiers à « souffrir » de ce type de spam. Le contenu du spam étant essentiellement fait de liens hypertextes, les sites destinataires voyaient leur rang augmenter du fait de ce très grand nombre de liens pointant vers eux.

La solution trouvée consista (en complément du déploiement de filtres spécifiques par les plateformes de création et/ou d'hébergement de blogs) à imposer l'attribut HTML `rel="nofollow"` qui, placé à l'intérieur d'une balise de lien, indique aux moteurs de ne pas suivre ledit lien et de ne pas en tenir compte dans leur algorithme de classement. En janvier 2005, les trois géants de la recherche d'information (Google, Yahoo! et Microsoft) se mettent d'accord pour systématiser

¹⁰¹ Pour une étude plus détaillée du phénomène, voir ce billet de Jean Véronis : <http://aixtal.blogspot.com/2005/09/web-google-blogger-et-le-splog.html>

cette procédure dans leur indexation des blogs¹⁰². Il ne manque plus alors que l'accord des plateformes de création elles-mêmes pour automatiser la mise en place de cette balise dans le cadre du dépôt de commentaires. Ce qui fut fait quelques semaines après par les principaux acteurs de ce marché (Live Journal, Six Apart, Blogger, MSN Spaces, etc.). Ainsi, désormais et par défaut, l'attribut `rel="nofollow"` est systématiquement placé dans le code HTML des commentaires, de manière transparente pour l'utilisateur et l'auteur du blog.

Si elle n'empêche pas le spam en lui-même, cette procédure diminue considérablement son impact (et donc son intérêt) sur le positionnement des sites dans les moteurs de recherche.

Les classements : influence ou affluence ? Autorité ou notoriété ?

Dès 2005, le premier outil à proposer un classement thématique et actualisé en permanence fut PubSub¹⁰³. Le site est aujourd'hui fermé mais, dans l'intervalle, nombre de classements de blogs ont fait leur apparition. Les deux principaux sont ceux de Wikio¹⁰⁴ – qui propose un classement mensuel et francophone – et celui d'Alianzo¹⁰⁵, à visée internationale, qui propose un classement par pays, par langue et par thématique.

Ces classements offrent au chercheur d'information un vivier de blogs « incontournables », qui peuvent constituer un bon point de départ pour certaines recherches. À la différence de ce que proposait à l'époque le service PubSub, il n'est cependant pas possible pour l'instant de récupérer en une seule fois l'ensemble des blogs de telle ou telle catégorie pour les ajouter dans son agrégateur¹⁰⁶.

Attention cependant à ne pas accorder trop de place à ces classements. Ils ne permettent de mesurer que la consultation d'un blog et sa notoriété. En aucun cas cependant la notion d'autorité n'est présente. L'immense majorité des blogs intéressants et dont l'auteur « fait autorité » sur son sujet (qu'il soit universitaire, journaliste, chercheur ou « simple » professionnel) ne sont pas présents dans ces « top 50 » de la blogosphère. Il est en revanche possible de localiser beaucoup de ces « appelés » en explorant systématiquement la blogoliste de ces quelques « élus ».

102 Les trois déclarations de Microsoft, Yahoo! et Google sont disponibles aux adresses :

http://blogs.msdn.com/msnsearch/archive/2005/01/18/nofollow_tags.aspx

www.ysearchblog.com/archives/000069.html

www.google.com/googlegblog/2005/01/preventing-comment-spam.html

103 www.pubsub.com/

104 www.wikio.fr/blogs/top

105 www.alianza.com/en/top-blogs/

106 Le format de fichier permettant de réaliser cette manipulation s'appelle OPML.

◆ **Exemples et usages**

De manière générale, pour une veille efficace, on s'efforcera de localiser et d'identifier les principaux blogs « d'audience » (les plus lus ou vus) et « de référence » (dans lesquels la valeur ajoutée de l'information présentée tient à la qualification de son auteur)¹⁰⁷. Ce « cœur de cible » une fois identifié, on le complètera utilement en identifiant un méta-blog sur le sujet ainsi que des blogs qui traitent de la thématique recherchée non pas de manière centrale mais davantage « à la marge » (par exemple dans une de leurs rubriques). Ces blogs périphériques permettront d'élargir le corpus et d'extérioriser un peu la recherche d'informations.

Zoom sur la blogosphère scientifique

Par « blogosphère scientifique », nous désignons l'ensemble des blogs se rapprochant des thématiques de l'IST (information scientifique et technique), aussi bien dans l'aspect recherche que dans l'aspect valorisation et diffusion.

Si ce domaine est très largement représenté au sein de la blogosphère, sa composition n'est cependant pas uniforme selon les disciplines. Il existe certains domaines plus développés que d'autres. Ainsi pour celui du droit on parle d'une « blawgosphère » tant les blogs de juristes, d'avocats, d'enseignants en droit sont nombreux.

La blogosphère « scientifique et technique » peut être découpée en plusieurs parties (figure 25).

La blogosphère institutionnelle

Cette blogosphère « institutionnelle » regroupe les blogs collectifs d'universités, de laboratoires ou d'équipes de recherche, de formations ou de cursus. Elle n'apporte pas de réelle valeur ajoutée et stigmatise simplement le glissement qui se fait depuis les anciennes pages personnelles/institutionnelles vers ces nouvelles plateformes éditoriales que sont les blogs. Il s'agit majoritairement de blogs créés dans une logique de communication institutionnelle.

Quelques exemples :

- blog du Master 2 « Droit des collectivités territoriales » de l'Université d'Évry¹⁰⁸
- blog de la présidence de l'Université de Nice¹⁰⁹
- blog de Richard Descoings, directeur de Sciences-Po Paris¹¹⁰

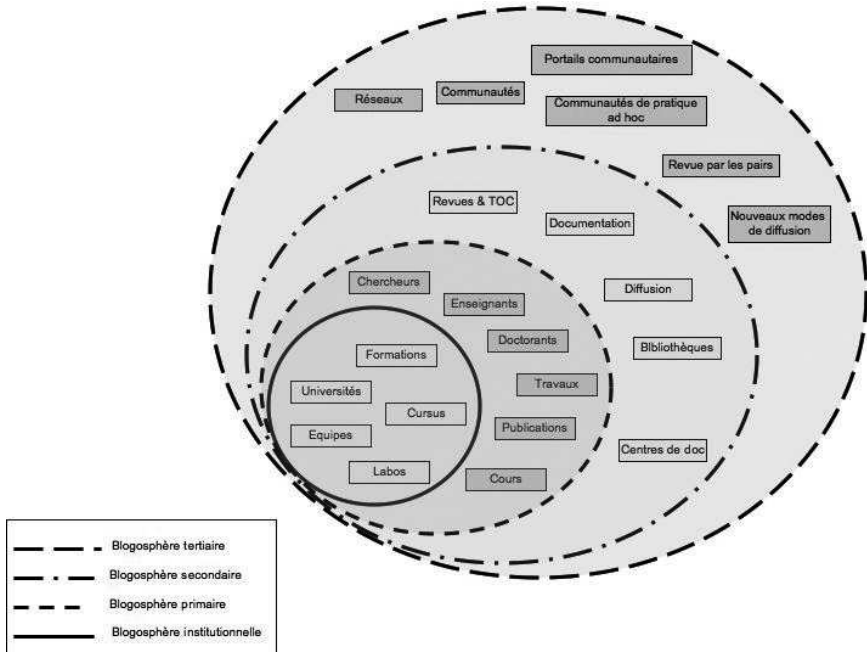
107 Cf. les trois notions déjà soulignées de notoriété, d'autorité, de régularité (dans l'alimentation du blog), auxquelles on peut ajouter l'ancienneté (quantifiable en consultant la rubrique « Archives » du blog).

108 <http://mastercollectivitesterritoriales.over-blog.com/>

109 <http://portail.unice.fr/blog/presidence/>

110 <http://richard-descoings.net/>

FIGURE 25 – UNE TYPOLOGIE DE LA BLOGOSPHERE SCIENTIFIQUE



- blog de Bernard Rentier, recteur de l'Université de Liège¹¹¹. C'est l'exception qui confirme la règle : ce blog se distingue par l'engagement et les prises de position de son auteur, notamment en faveur du mouvement de l'Open Access.

La blogosphère « primaire »

Cette blogosphère « primaire » (au sens où l'on parle de « document primaire ») regroupe les blogs individuels d'enseignants, de chercheurs, de doctorants. Elle permet à ses acteurs de faire état de leurs publications, de leurs travaux, de leurs réflexions, ainsi que de revenir ou de mettre en avant tel ou tel projet de recherche et ses différents états d'avancement. On y trouve également de nombreux blogs servant à la mise en ligne de supports de formation ou de cours calibrés dans une logique de formation à distance, le plus souvent à l'initiative des individus eux-mêmes, mais parfois également à l'initiative de leur institution ou organisme de rattachement.

Quelques exemples :

- blog de chercheur : Jean Véronis, professeur de linguistique et d'informatique¹¹²
- blog de chercheur : Jean-Michel Salaün, directeur de l'École de bibliothéconomie et des sciences de l'information (EBSI) au Canada¹¹³

¹¹¹ <http://recteur.blogs.ulg.ac.be/>

¹¹² <http://aixtal.blogspot.com/>

¹¹³ <http://blogues.ebsi.umontreal.ca/jms/>

- actualités de la recherche en histoire visuelle : blog du Laboratoire d'histoire visuelle contemporaine de l'École des hautes études en sciences sociales (EHESS), alimenté par André Gunthert, maître de conférences¹¹⁴
- blogs de la faculté de Droit de Harvard¹¹⁵
- blog de cours en ligne : Veille stratégique¹¹⁶
- blog de cours en ligne : DUT Information et communication de l'IUT de La Roche-sur-Yon (option métiers du livre)¹¹⁷

La blogosphère « secondaire »

Cette blogosphère secondaire (au sens où l'on parle d'un document secondaire) regroupe les blogs individuels de documentalistes et/ou de bibliothécaires spécialisés, les blogs collectifs de centres ou d'organismes à vocation documentaire, ainsi que les blogs ou sites permettant de diffuser les tables des matières d'un grand nombre de revues grâce à des fils RSS. Elle regorge souvent de pépites documentaires à très forte valeur ajoutée. Elle est incontournable pour qui veut effectuer une veille documentaire.

Quelques exemples :

- libre accès à l'information scientifique et technique (INIST)¹¹⁸
- Prosper : actualités en IST (INIST)¹¹⁹
- écrans de veille en éducation : blog de l'Institut national de la recherche pédagogique (INRP)¹²⁰
- Zetoc : service de la British Library donnant accès aux sommaires de 20 000 journaux scientifiques et 16 000 actes de conférences depuis 1993¹²¹

La blogosphère « tertiaire »

Cette blogosphère « sociale » ou « tertiaire » est la plus récente. Elle marque une étape significative dans ce que certains appellent la « science 2.0 », c'est-à-dire de nouvelles manières, encore expérimentales, de diffuser et d'évaluer les résultats de la science, ainsi que de constituer de nouveaux laboratoires scientifiques.

Elle comprend des communautés de pratique ou d'intérêt *ad hoc*, comme par exemple l'ensemble des blogs du monde des bibliothèques et de la documentation, ou bien encore la « blawgosphère » (blogosphère juridique), et surtout des portails communautaires exploitant toute la richesse des fonctionnalités des blogs et de l'agrégation de fils RSS.

114 www.arhiv.lhivic.org/index.php/

115 <http://blogs.law.harvard.edu/>

116 www.glsreseaux.com/veille_strategique/

117 <http://oecoursiut.wordpress.com>

118 <http://openaccess.inist.fr/>

119 <http://prosper.inist.fr/>

120 <http://prosper.inist.fr/>

121 <http://zetoc.mimas.ac.uk>

Quelques exemples

- Postgenomic¹²². Ce site collecte et affiche nombre de blogs « scientifiques » dans les domaines de la chimie, de la bio-informatique, de la médecine, des sciences de la vie et de la terre, des neurosciences et de la physique. Il propose ensuite un certain nombre de services permettant notamment de dégager des tendances, de regarder quels sont les articles les plus cités dans les billets des blogs, de pointer les *reviews* d'un article dans les mêmes billets, de lister sous forme de « zeitgeist » les revues les plus citées par les blogueurs du domaine, etc.
- Researchblogging¹²³. Ce service propose aux blogueurs (scientifiques) de s'inscrire et d'inclure une ligne de code dans leurs billets à partir du moment où ces derniers s'apparentent au processus de revue par les pairs, c'est-à-dire où ils rendent compte, en la commentant, d'une publication scientifique. Le service Researchblogging collecte alors ces billets pour les afficher dans des catégories thématiques correspondant aux disciplines scientifiques des articles « repérés ». Soit une réelle valeur ajoutée pour « filtrer » tous les autres billets « moins » scientifiques. Au 31 janvier 2008, Researchblogging ne s'appuie « que » sur cent cinquante blogs scientifiques, dont la plupart sont eux-mêmes issus du site portail Scienceblogs¹²⁴ qui en rassemble déjà près de soixante-dix.
- Edufeeds¹²⁵. Ce site est un portail au sens classique du terme, c'est-à-dire qu'il rassemble de manière thématique des ressources validées et organisées en arborescence. Cet annuaire recense un nombre conséquent de liens vers les pratiques universitaires en termes de blogs, de RSS et de *podcast* (balado-diffusion). Le classement est thématique (usages en : Bibliothèques, Journaux, Ressources humaines, Étudiants, Centres et laboratoires de recherche, etc.). La liste des blogs de chaque catégorie est téléchargeable sous la forme d'un fichier OPML.

Identifier les blogs leaders sur des marchés de niche

Le pouvoir de prescription des blogs est aujourd'hui reconnu. Ils sont entrés dans nombre de stratégies marketing habituelles pour accompagner la sortie d'un produit et alimenter le « buzz », le bouche à oreille. Il peut donc être intéressant de repérer les quelques blogs « influents » sur tel ou tel secteur.

En complément des méthodes et critères déjà rappelés (sur l'utilisation des moteurs de recherche et l'exploration systématique des fonctionnalités

122 www.postgenomic.com

123 www.researchblogging.org

124 www.scienceblogs.com

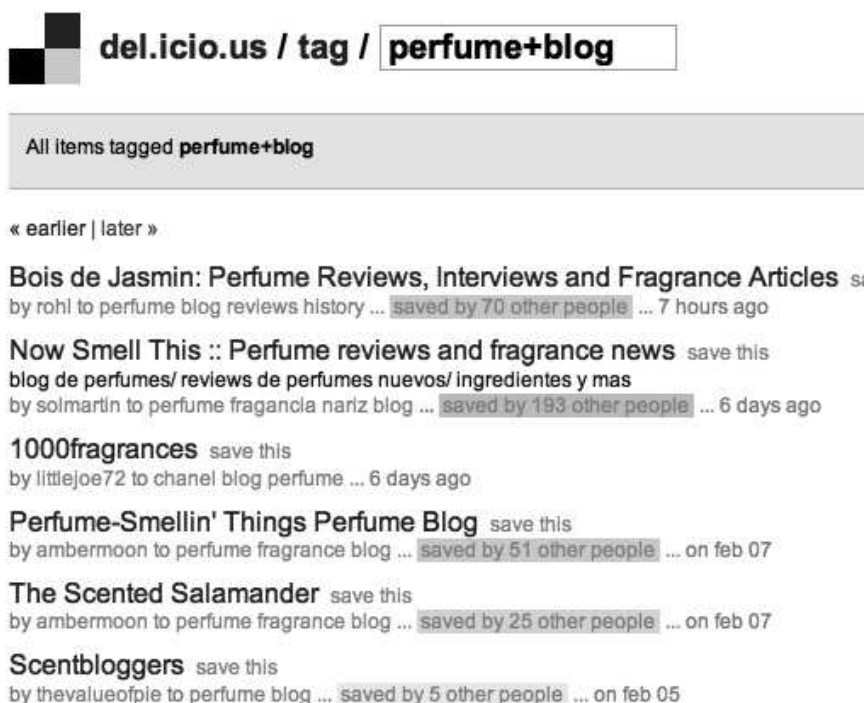
125 <http://directory.edufeds.com>

propres au blog – blogoliste par exemple), deux outils s'avèrent particulièrement adaptés pour ce genre de démarche¹²⁶.

- Blog Finder de Technorati¹²⁷. Le Blog Finder de Technorati permet de filtrer les blogs selon des catégories thématiques, d'affiner ensuite les résultats selon la pertinence, l'autorité (il s'agit en fait d'une mesure de notoriété, le tri s'effectuant selon le nombre de liens pointant vers un blog donné) ou la fraîcheur (billets les plus récents), et enfin de télécharger le fichier OPML des dix premiers résultats affichés.

- Tags de Del.icio.us. L'outil de signets partagés Del.icio.us peut également être exploité dans une même logique. En saisissant une adresse de ce type : <http://del.icio.us/tag/mot-clé+blog>, on fait rapidement remonter les blogs traitant de la thématique, du sujet ou du produit que l'on cherche à cibler. Il suffit alors d'isoler les blogs les plus « partagés / bookmarkés ».

FIGURE 26 – IDENTIFIER DES BLOGS « LEADERS » SUR DEL.ICIO.US



126 Le billet en date du 15 février 2008 du blog ReadWriteWeb présente une liste plus complète de ces outils et des avantages et inconvénients de chacun d'entre eux. www.readriteweb.com/archives/identify_top_blogs.php

127 <http://technorati.com/blogs/directory/>

Conclusion

« J'ai dix ans »... de plus



À l'heure où paraît cet ouvrage, les blogs, ce « premier format éditorial né du web », viennent de fêter leur dixième anniversaire¹²⁸. Leur histoire, à l'image des nombreuses aventures qui émaillent l'histoire du réseau Internet, a d'abord pris naissance aux États-Unis avant de devenir aujourd'hui un phénomène mondial.

La gratuité des plateformes, la simplicité des usages, la richesse des fonctionnalités proposées, l'attrait et l'importance croissante des logiques communautaires sont autant de facteurs qui permettent d'expliquer ce formidable succès. Du journal intime de l'adolescent au bloc-note du chercheur en passant par le relevé de terrain du commercial ou du salarié, les blogs touchent aujourd'hui l'ensemble des professions, des statuts et des couches sociales.

◆ De quoi seront faits les blogs dans dix ans ?

En 2004, l'un des tous premiers ouvrages parus sur le phénomène des blogs indiquait : « Au risque d'étonner, on peut soutenir qu'à un horizon de quelques années – disons cinq ans – les blogs auront disparu. Non pas que les gens vont arrêter de bloguer, ou que le blogging n'aura constitué qu'un passe-temps temporaire pour quelques passionnés d'Internet. Bien au contraire, les blogs en soi vont disparaître parce qu'ils vont se banaliser, se généraliser, et se fondre avec les outils que nous utilisons, et les usages qui en émergent, pour jouer un rôle majeur au cœur de notre utilisation d'Internet.¹²⁹ »

Plusieurs points de cette analyse peuvent être aujourd'hui confirmés.

◆ La généralisation des blogs est effective

Et ce dans les trois sphères de notre environnement informationnel : la sphère publique pour les blogs rattachés à une activité professionnelle, à un loisir ou

¹²⁸ Depuis l'invention du terme en 1997, et les premières plateformes grand public apparaissant dans les années 1998-1999.

¹²⁹ Cyril Fiévet et Emily Turrettini, *Blog Story*, page 238 (voir la référence page 63).

à une passion ; la sphère privée pour les blogs utilisés dans le cadre d'un réseau fermé sur l'extérieur (intranets) ; la sphère intime pour les blogs qui laissent voir, consciemment ou inconsciemment, une part plus ou moins grande de l'intimité de son auteur et parfois de ses proches.

D'autre part, l'essor du Web dit « 2.0 », avec la part et la place de plus en plus large qui sont accordées aux points de vue et opinions des usagers sur tous les supports et dans tous les domaines¹³⁰, permet rétrospectivement d'accorder aux blogs un rôle de pionnier dans cette irruption de l'utilisateur lambda au cœur des contenus web.

◆ Leur banalisation est actée

Un point de rupture significatif, allant dans le sens d'une banalisation des blogs, est celui qui vient d'être franchi par la plupart des plateformes de création et d'hébergement, qui autorisent désormais la création et la gestion distinctes de « pages » (au sens habituel de l'unité minimale de publication d'un site web) et de billets. On a déjà pu observer que la plupart des « pages personnelles » qui se créent aujourd'hui sur le Net ne sont plus, loin s'en faut, des sites ou pages web « classiques », mais des blogs permettant de mettre en valeur un CV ou une passion.

La seule limite des blogs, qui freinait leur adoption par des sites à vocation plus institutionnelle ou associative, était jusqu'ici la contrainte de l'affichage et de la gestion rétro-chronologiques des billets¹³¹. La possibilité aujourd'hui offerte de jouer sur les deux tableaux (billets ET pages) vient combler ce manque. Grâce à la création de pages sur un blog, il est désormais possible de faire fonctionner toute la richesse d'une organisation arborescente en s'affranchissant des contraintes d'affichage et d'organisation liées aux « billets ». Cette nouvelle possibilité offerte implique des changements qui vont bien au-delà du simple aspect technique.

De tout cela on peut conclure que les blogs, en se multipliant, en se banalisant, en investissant toutes les pratiques professionnelles, sociales et relationnelles, ne forment pas simplement un immense réseau hypertextuel de contenus se répondant entre eux, mais également une foule de réseaux sociaux de densités variables¹³².

130 Phénomène du Crowdsourcing (sagesse des foules) et de l'UGC (User Generated Content : contenu généré par l'utilisateur).

131 Même si certaines plateformes, pour contourner ce facteur bloquant, proposaient déjà de placer un billet systématiquement en tête du blog, pour qu'il fasse office de « page d'accueil ».

132 Jan Schmidt, *Blogging practices...* (voir la référence page 64).

Annexes



◆ Références pour approfondir...

Tous les liens ont été vérifiés le 14 mai 2008.

Ouvrages

Christophe ASSELIN, Véronique MESGUICH, *Le web 2.0 pour la veille et la recherche d'information*, Livre blanc édité par la société Digimind, juin 2007, www.digimind.fr/services/white_paper.htm

Serge COURRIER, *Utiliser les fils RSS et Atom*, ADBS Éditions, 2007 (Coll. « L'essentiel sur... »).

Cyril FIEVET, Emily TURRETTINI, *Blog Story*, Paris, Eyrolles, 2004

Articles

Rebecca BLOOD, « Weblogs: A History and Perspective », Rebecca's Pocket. Mise en ligne : 7 septembre 2000, mise à jour : 25 octobre 2006. www.rebeccablood.net/essays/weblog_history.html. Traduction française sur www.elanceur.org/Traductions/RebeccaBloodWeblogsHisto.html

Sylvie DALBIN, « Comment utiliser les blogues pour la diffusion de dossiers documentaires ». http://dossierdoc.typepad.com/dossierdoc/2004/12/comment_utilise.html

Olivier ERTZSCHEID, « Weblogs : demain... la veille » in *Bases*, septembre 2005, n° 219. www.bases-publications.com/revues/bases/e-docs/00/00/02/A2/document_article.phtml

Cyril FIEVET, Gilles KLEIN, « L'ABC du blog. ». www.lemondedublog.com/abc/

Martin ROELL, « Distributed KM – Improving Knowledge Workers' Productivity and Organisational Knowledge Sharing with Weblog-based Personal Publishing ». www.roell.net/publikationen/distributedkm.shtml

Jan SCHMIDT. « Blogging practices: An analytical framework ». *Journal of Computer-Mediated Communication*, vol. 12, n° 4, article 13. <http://jcmc.indiana.edu/vol12/issue4/schmidt.html>

Autres sites

OUR BOLD HERO, « Masterlist » : liste d'articles scientifiques consacrés au phénomène des blogs dans différents contextes (sociologique, technologique, entrepreneurial, etc.). www.ourboldhero.com/research/masterlist.html

KAIROS NEWS, « Weblog Webliography » : présente près de 200 articles (de presse ou billets de blogs et quelques articles d'universitaires) consacrés aux différents aspects des blogs, principalement autour de la thématique de l'enseignement. <http://kairosnews.org/blogbib>

BLOGBIB, « Blogbib » : ce blog (qui n'est plus alimenté depuis janvier 2007) présente une sélection de ressources thématiques autour des blogs (définitions, historique, ouvrages dédiés, articles, études, présentations). <http://blog-bib.blogspot.com/>

URFIST de Paris (ressource pédagogique), « Les Blogs », créé le 19 octobre 2005. Mise à jour : 11 octobre 2006. www.ext.upmc.fr/urfist/rss/blog.html

◆ Compléments en ligne

Deux autres documents annexes sont accessibles en ligne sur le site de l'ADBS (www.adbs.fr), à partir de la notice de présentation de cet ouvrage dans la rubrique « Accéder à la documentation professionnelle » :

Douze conseils pour devenir un bon blogueur

Ils ont dit... : les avis d'une douzaine de personnalités sur le phénomène des blogs

Impression : Compédit Beauregard s.a.
61600 La Ferté-Macé

Dépôt légal : 2^e trimestre 2008

N° d'Imprimeur : 6756